Predicting Firm Profits: From Fama-MacBeth to Gradient Boosting*

Murray Z. Frank Keer Yang

August 28, 2021

Abstract

This paper studies the predictability of firm profits using Fama-MacBeth regressions and gradient boosting. Gradient boosting can use more relevant factors and it predicts better. Profits are more predictable at firms that are large, investment grade, low R&D, low market-to-book, low cash flow volatility. Effects on financing decisions, and cross-section of stock returns are studied. During recessions profits are less predictable - particularly non-investment grade firms. Both algorithms produce estimates like those interpreted in the literature as evidence of excessive human optimism during booms and excessive pessimism during recessions.

Key Words: Expected profit, Fama-MacBeth, gradient boosting, firm financing decisions, cross-section of stock returns, behavioral finance
JEL Classification: G17, G32, G40

*Respective affiliations are Murray Frank, murra280@umn.edu, University of Minnesota, Minneapolis, MN; and SAIF, Shanghai, China. Keer Yang, yang5427@umn.edu, University of Minnesota, Minneapolis, MN. Murray Frank thanks Piper Jaffray for financial support. We thank Vojislav Maksimovic and Jeremiah Green for helpful comments. The usual disclaimer applies.

1 Introduction

In theory firms try to maximize expected profits. Despite the central role of expected profits, it is not entirely clear how best to measure those expectations. In this paper we study the efficacy of two fundamental methods of predicting firm profits: traditional estimation using Fama and MacBeth (1973) regressions, and estimation using gradient boosting (Friedman, 2002; Chen and Guestrin, 2016). We find that gradient boosting provides better forecasts as it permits a much larger set of factors to play a predictive role. The 149 potential factors we use include firm attributes and macro factors from Fama and French (2006); Welch and Goyal (2008); Frank and Goyal (2009); Frank and Yang (2019); Gu et al. (2020). Having established the statistical efficacy of the prediction methods, we then apply the predictions from both methods in three familiar finance settings to verify the practical efficacy of the methods. We study a corporate finance application (flows of debt and equity), an asset pricing application (cross-section of stock returns) and a behavioral finance application (predicting prediction mistakes).

This paper contributes four main results to our understanding of expected profits. 1) Gradient boosting generates higher quality firm profit predictions, and those predictions align with many observed firms decisions in a reasonable manner. 2) The distinction between current profit and current expectation of future profit, is important for firm financing decisions. Both matter but in distinct ways. Current profits are not a fully satisfactory proxy for expected profits. 3) Gradient boosting profit predictions act similarly but somewhat more strongly than gross profits (Novy-Marx, 2013) when used for the cross-section of stock returns. So it may offer an improved proxy in some standard asset pricing applications. 4) Neither Fama-MacBeth nor gradient boosting generate profit predictions that satisfy rational expectations. We apply an econometric test from Bordalo et al. (2021). They used the estimated coefficients as evidence that stock analysts overreact, making predictions that are too optimistic in booms and too pessimistic in recessions – 'diagnostic expectations'. We find that the profit predictions from Fama-MacBeth regressions and from the gradient boosting algorithm generate the same basic patterns of estimated coefficients as Bordalo et al. (2021) found for human stock analysts. Consider these four results in turn.

First, the gradient boosting approach (denoted GBRT), due to Friedman (2002) produces better firm profit predictions than does the Fama and MacBeth (1973) approach (denoted FM). This is true both in-sample and out-of-sample. It is primarily due to the ability to include many more factors without over-fitting the data.

Many key effects of predicted profits hold under both approaches to prediction. Comparing firms, we find that large firms and investment grade firm profits are more predictable than average firms. Firms with high R&D, market-to-book, and cash flow volatility have less predictable profits than average firms. Among publicly traded firms that exit, unprofitable firms tend to be liquidated or bankrupt; while profitable firms tend to be involved in an acquisition, a merger, an LBO, or to become a private firm. During the financial crisis of 2007-2009 and during NBER recessions, firm profits become less predictable. The reduced predictability during bad times affects average firms much more than it affects investment grade firms.

Second, firm profits are often used as a control variable rather than being the main focus of a paper, see Mitton (forthcoming). In such settings current profits are often used as an easy proxy for expected future profits. Our results show that actual current profits are not an entirely satisfactory proxy for expected profits. Firm net debt and net equity decisions are affected quite differently by the two, as predicted by the model in Frank and Sanati (forthcoming).

Third, the gradient boosting based profit predictions can be viewed as an alternative to the familiar profit proxies when studying stock returns (Fama and French, 2006; Aharoni et al., 2013; Novy-Marx, 2013). The direction of the effects are the same and even the estimated coefficients are close, but the gradient boosting based proxy seems to have a somewhat stronger statistical effect in the cross-section of stock returns. We do not interpret this use of gradient boosting profit predictions as a 'new factor' (Harvey et al., 2016). Instead, we view it as an alternative proxy for an established result – that expected profit is important for stock returns (Fama and French, 2006; Aharoni et al., 2013; Novy-Marx, 2013).

Fourth, the predictions generated by gradient boosting are very good, but not good enough to satisfy the restrictions of rational expectations equilibrium when examined carefully. Prediction mistakes are forecastable using linear regressions using past profits or past prediction mistakes as factors. This kind of predictable prediction mistakes are described in the literature as evidence of excessive human optimism during booms and excessive pessimism during recessions, see Bordalo et al. (2021). Of course, by construction the algorithms treat all observations equally at the start, and neither algorithm is human.

1.1 Related literature

We use the terms expected profits and predicted profits as synonyms because nothing that we do rests on there being an important distinction. The profit prediction problem itself is high dimensional with potentially important nonlinearities. We interpret the gradient boosting predictions as feasible proxies for the otherwise unobservable expectations. A similar idea has been developed in more detail by Nagel (2021).

There are several machine learning methods that could be adopted. Gradient boosting, random forest, and deep learning are all high profile machine learning algorithms with distinct strengths and weaknesses, see (Hastie et al., 2009; Efron and Hastie, 2016). Erel et al. (2021) successfully used several machine learning algorithms to study the firm's selection of directors.

In our view gradient boosting provides a good balancing of attributes. Random forest is the most automatic of the three algorithms. Deep learning requires the most effort to refine performance. Our use of gradient boosting also reflects our past experience that the algorithm performs well when applied to firm data. We anticipate that this algorithm will also prove helpful for other corporate finance studies. However, it should be noted that performing well in practice, is not the same thing as optimal in an absolute sense. Random forest for example is occasionally described as optimal (van Binsbergen et al., 2020). Efron and Hastie (2016) point out that "Random forests are somewhat more automatic than boosting, but can also suffer a small performance hit as a consequence."

Deep learning models perform extremely well in some applications. However, they also often turn out to be underspecified with problematic hold-out performance concerns more often than sometimes recognized. This issue is covered at length by D'Amour et al. (2020).

Ensemble methods often perform better than individual algorithms, see Hastie et al. (2009). This is well known. So we do not claim that gradient boosting is statistically optimal in an unrestricted sense. In particular, our finding of linear predictability of prediction errors appears to be a reflection of this well-known advantage of ensembles.

The behavioral finance literature is often interested in showing the failures of rational expectations. A particularly interesting version is provided by Bordalo et al. (2021). They show that stock analyst forecasts contain predictable mistakes that are too optimistic when things are good and too pessimistic when things are bad. They use that evidence to motivate a structural model of diagnostic expectations. Instead of testing human predictions, we apply their test to the profit predictions generated by Fama-MacBeth and by gradient boosting. We find that the algorithms generate the same patterns of predictable prediction mistakes as they found for IBES stock analysts. These algorithms fail rational expectations in much the same way as humans.

Nagel (2021) suggests that machine learning methods might offer a reasonable as-if model of purely human forecasting in high dimensional environments. Our test results seem broadly supportive of his idea. Various machine learning algorithms have distinct strengths and weaknesses. So in future research it might be of interest to examine which machine learning algorithms produce results that are most similar to human financial decisions.

In corporate finance, flows of debt and equity to firms are commonly interpreted through the lens of the tradeoff theory and the role of profits has been prominent, see Myers (1984); Fama and French (2002); Danis et al. (2014); Frank and Goyal (2015); Eckbo and Kisser (2021); Ai et al. (2021). We contribute to that literature by providing evidence that supports the model in Frank and Sanati (forthcoming). That model is based on a tradeoff of tax benefits of debt against the need for collateral when issuing debt. The distinction between current profits and expected profit plays a central role in their model. We find that as predicted, firms with high current profits tend to issue debt and repurchase equity. Firms with high expected profits tend to issue equity and repurchase debt.

Our method of estimating expected profits may also contribute a useful alternative proxy to the literature on stock returns (Aharoni et al., 2013; Novy-Marx, 2013; Fama and French, 2015). It is commonly thought that profits plays an important role for stock return. Fama and French (2006) used income before extraordinary items. Novy-Marx (2013) used gross profits. We use the same testing methods as those studies. Going beyond those studies, we find that gradient boosting estimated expected profits provides results that are similar but empirically somewhat stronger. Accordingly our approach can provide an alternative proxy for the role of expected profits. The ability of gradient boosting to impound a larger number of factors in the expectation seems helpful. Our contribution here is an alternative proxy for known results, not a new factor; nor do we enter into the debates over the specific set of best factors (Harvey et al., 2016; Hou et al., 2019).

2 Data

The initial firm data is from CRSP/Compustat Merged annual data extracted from WRDS, covering the years 1950-2019. Data is dropped if it is prior to 1964, for firms not based in the USA, firms in the Finance, Insurance and Real Estate industries (SIC codes from 6000 to 6999), missing key data items, or has negative book equity. The factor data is winsorized on an annual basis at the the 0.5% and 99.5% levels. The data extraction and cleaning steps are described in detail by Table 1 showing the impacts on the number of observations. The result is 121,401 firm/year observations.

The accounting explanatory variables for date t are for fiscal year that ends in calendar year t. Consistent with Aharoni et al. (2013) calculations are on a per firm basis, rather than the per stock basis used by Fama and French (2006). This is intended to address concerns about the impact of changes to the number of shares from year to year.

The variable to be predicted is profits. But the concept of profits does not exactly match the standard accounting data. As a result a variety of alternative measures have been used in different papers, see Mitton (forthcoming). Our tabulated results are for

Table 1: Data cleaning

1	start	CRSP/COMPUSTAT Merged Foundamental Annual data file (1950 - 2019, Consolidation Level "C")	326,248
2	keep	if datafmt = "STD"	0
3	keep	if indfmt = "INDL"	0
4	keep	if fic = "USA"	- 27,541
5	drop	if sic $>= 6000$ & sic $<= 6999$	- 84,666
6	drop	missing Fama French items, profitability, and total assets at t	-12,225
7	drop	negative book equity	- 8,065
8	drop	gvkey fyear duplicates	-2,205 (191546)
9	drop	missing total assets at $t-1$	-19,911
10	drop	if total assets less than $5 million or book equity less than $ 2.5 million	- 9,900
11	keep	if year(t) $<= 2014$ and year(t) $>= 1964$	- 12,496
11	winsorize	variables are winsorized when outside the 0.5 and 99.5 percentage each year	
12	training sample	year(t) $<= 2014$ and year(t) $>= 1964$ including profit information at t+1 (year 2015)	146,239
13	testing sample	year(t) $<= 2014$ and year(t) $>= 1975$ including profit information at t+1 (year 2015)	132,612
14	analysis sample	11,210 observations have missing total assets at t+1	121,401
15	analysis sample	1,447 firms have only one observation	119,955

operating profit. It is defined as

$$\pi_{i,t} = (Sales_{i,t} - COGS_{i,t} - SGA_{i,t})/AT_{i,t}. \tag{1}$$

We refer to $\pi_{i,t}$ interchangeably as 'profit' or 'profitability'.

The tests were also carried out using gross profit (Novy-Marx, 2013), and using income (Fama and French, 2006) as the dependent variable. In untabulated results, we also tried a number of further profit proxies. The inferences from our tests are very similar across profit measures. In earlier drafts we also tabulated the results from gross profits and from income. But for ease of reading these are no longer tabulated.

Theory does not specify the factors to be used when predicting profits. Presumably, anything that helps to predict the future profits should be included. As a practical matter, when using linear regressions or Fama-MacBeth, multicollinearity sharply limits the number of factors that can be included. A major potential advantage of using the gradient boosting is the ability to include a large number of candidate factors.

We studied two sets of explanatory factors, X. The first version of X_t is the list of factors used by Fama and French (2006), denoted "FF06". This provides an important foundation. The second version of X_t is a set of 149 factors from Fama and French (2006), Welch and Goyal (2008), Aharoni et al. (2013), Frank and Yang (2019), and Gu et al. (2020), denoted "All". All includes a wide range of firm level accounting measures as well as many macroeconomic factors. These are listed in the Appendix Table A1. No attempt was made to delete factors that are similar to each other. That task is left to the algorithms.

Time affects the analysis in several ways. First, variables are updated at different frequencies. Most of our analysis is at annual frequency and we use the most recent quarter or month as the factor. Second, predictions are made for 1976 to 2015 using information from 1964 to 2014. To make sure that there is enough information to correctly estimate the coefficients, we start to predict profits in year 1975 using information from year 1964 to 1974.

For in-sample estimation all data is used together. For out-of-sample estimation only data from prior dates is included in the rolling estimation. When calculating expected profit $\hat{\pi}_{i,\tau+1}$, only information until year τ are used. We do not use cross-sectional coefficients estimated using information at $\tau + 1$ or further in the future in order to avoid the look-ahead bias. We call the rolling sample estimates 'out-of-sample' for simplicity. Some authors prefer the term pseudo-out-of-sample, since it uses data that is actually already in the past when a study is conducted.

3 Methods to make predictions

This section explains the classic Fama-MacBeth method and the modern gradient boosting method that are used to produce the profit predictions. For Fama-MacBeth see Fama and MacBeth (1973), Petersen (2009) and Campbell (2017). For gradient boosting see Hastie et al. (2009), Chen and Guestrin (2016) and Efron and Hastie (2016). The reported Fama-MacBeth estimation results were implemented using Stata 14. The reported gradient boosting estimation results were implemented using scikit-learn library (version 0.24.2), see https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html. Earlier drafts of this paper implemented gradient boosting using XGBoost which was also used by Erel et al. (2021). Both software libraries generate extremely similar results.

3.1 Fama-MacBeth

Following Fama and French (2006) and Aharoni et al. (2013) we estimate a series of regressions to predict future profit.

$$\pi_{t+1} = \lambda_0 + \lambda_1 X_t + \varepsilon_{it} \tag{2}$$

where π_{t+1} is profit during the fiscal year that ends in calendar year $t+1$. X_t is the list of market and accounting factors. Versions using FF06 and using All are estimated.

Parameters are estimated on a rolling basis. The coefficients $\hat{\lambda}_0^t, \hat{\lambda}_1^t$ used to calculate predicted $\hat{\pi}_{t+1} = \hat{\lambda}_0^t + \hat{\lambda}_1^t X_t$, are the average slope coefficients from year-by-year cross-section profit regressions up until year t. When calculating expected profit $\hat{\pi}_{t+1}$, only information prior to year t are used. The data is from 1964 to 2015. The predictions are estimated for years 1975 to 2015, so each estimate has more than a decade of data. As time passes the training sample contains a gradually growing number of observations. For greater detail, see the algorithm 1 in the Appendix.

Algorithm 1 Fama and MacBeth (1973) Predictions

procedure FM($X_{i,t}, \pi_{i,t+1}$)▷ Where t - time year, i - firm, $X_{i,t}$ - predictors at time t, $\pi_{i,t+1}$ - profitability to predict
  for $1975 \leq T \leq 2014$ **do**
    for $1964 \leq t \leq T-1$ **do**
      run cross-sectional regression for all firms at time t
      $\pi_{i,t+1} = \lambda_{0,t} + \lambda_{1,t} X_{i,t} \varepsilon_{i,t}$
      get estimated coefficients $\{\hat{\lambda}_{0,t}, \hat{\lambda}_{0,t}\}$
    end for
    Compute average coefficients:
    $\hat{\lambda}_{0,T} = \sum_{t=1964}^{T-1} \hat{\lambda}_{0,t}/(T-1964)$
    $\hat{\lambda}_{1,T} = \sum_{t=1964}^{T-1} \hat{\lambda}_{1,t}/(T-1964)$
    Use information at time T $X_{i,t}$ to make prediction of profitability at $T+1$
    $\hat{\pi}_{i,T+1} = \hat{\lambda}_{0,T} + \hat{\lambda}_{1,T} X_{i,T}$
  end for
end procedure

3.2 Gradient boosted regression trees

Gradient boosting regression trees are a particularly prominent and empirically successful prediction method used in many applications. It starts with regression trees and then refines them iteratively by focusing on the errors in the previous iteration. This idea of refinement by focusing on error correction is known as boosting. Gradually an entire forest of trees is constructed. Then an average across the set of trees is used as the model's prediction.

Start with a regression tree. It is the basic building block for the method. A regression tree uses numerical cutoffs to assign an observation to a branch. In the simplest case there is a threshold number $\overline{x}$. If the observation has $x > \overline{x}$ the data is assigned to the upper branch. If the observation has $x \leq \overline{x}$ the data is assigned to the lower branch. Then within each branch there are subsequent partitions constructed. Eventually an entire set of branches are constructed. The final set is called a leaf, and it defines the set that the particular observation belongs to. A range of rules can be used to define the number of branches, their order of consideration and so on.

Figure 1: Tree Example

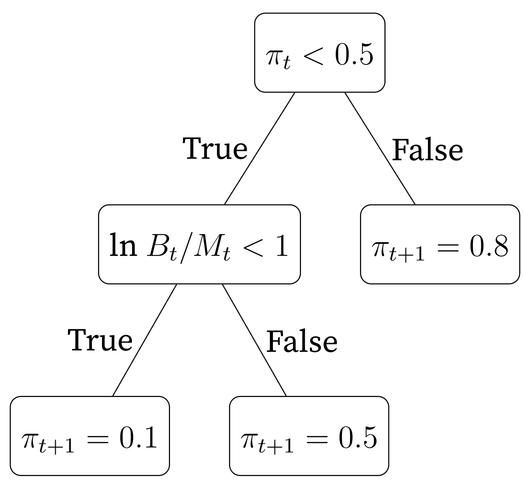


Figure 1 shows an example of a simple tree. It predicts expected profit using current profit, and book to market. If current profit is greater or equal to 0.5, the decision tree predicts the future profitability to be 0.8. If current profit is less than 0.5 and log value of book to market is less than 1, predicted future profit is 0.1. Otherwise, the predicted profit is 0.5. The tree in Figure 1 can also be represented as,

$$\hat{y} = h(x) = \sum_{m=1}^{3} c_m I\{(x_1, x_2) \in R_m\} \tag{3}$$

where R_m is the partition of the input variables, and c_m is the predicted value assigned to the terminal leaf.

In the example it takes at most two "branches" to reach the final terminal leaves. This is called the number of layers of the decision tree. The more layers the tree has, the more complex the model is. When a regression tree model is deep (more layers), the model tends to have low bias but larger variance. While when the model is shallow (fewer layers), the model becomes too simple, with low variance but large bias.

Regression trees are easy to interpret, but commonly do not predict well. Ensembles of trees have been found to predict better. Random forest Hastie et al. (2009) constructs many trees using independent bootstrap samples from the data to generate a forest. Boosting creates trees to improve on past prediction mistake instead of bootstrap samples from the original data. Boosting is more complex than bootstrapping, but it tends to improve final performance, see Efron and Hastie (2016).

Gradient boosting is the most widely adopted version of boosted forests. It starts by estimating decision trees with fixed shallow depth. Then it computes the residuals for the trees. At the next iteration more weight is devoted to the cases in which the model fit poorly. In the end an ensemble of trees are used to 'vote' on the appropriate results. This generally reduces the bias in a simple tree model while maintaining the low variance. The main drawback relative to a simple tree as in Figure 1, is that forests do not have such simple depictions that show how each variable affects the final result.

A more formal representation is,

$$\hat{y} = F_M(x) = \sum_{m=1}^{M} h_m(x) \tag{4}$$

where h_m is decision tree regressor with depth of d, and M is the number of trees in the forest. $F_M(x)$ is solved by using a greedy algorithm framework,

$$F_k(x) = F_{k-1}(x) + \gamma h_k \tag{5}$$

where γ is the learning rate. The learning rate shrinks the contribution of each additional tree. h_k is the newly added tree solved by minimize a loss function L given $F_{k-1}(x)$

$$h_k = \arg\min_h \sum_i (L(y_i, F_{k-1}(x_i) + \gamma h(x_i)) \tag{6}$$

There are three important hyperparameters in gradient boosting: the depth of the tree d (max_depth), the number of trees in the forest M (n_estimators), and learning rate γ (learning_rate). The default parameters have the following values: depth of the tree is 3, number of trees in the forest is 100, and learning rate is 0.1. We have systematically carried out the analysis with the default hyperparameters as well as with hyperparameters optimized using cross-validation. The results are very similar. Except where noted, the Tables use the default hyperparameters.

As in the Fama-MacBeth estimation, to predict profit at time $\tau + 1$, $\hat{\pi}_{\tau+1}$, we train the model using information before $\tau+1$, $\{X_{t-1}, \pi_t\}_{t\leq\tau}$. We then apply the model using factors available at time τ, X_τ.

Gradient boosting estimation was done using the software GradientBoostingRegressor from https://scikit-learn.org/stable/modules/ensemble.html, and using XGBoost from Chen and Guestrin (2016). The results are very similar. The reported results in the Tables use the algorithm GradientBoostingRegressor from scikit-learn.org. For greater detail see the algorithm 2.

4 Methods to evaluate predictions

There are many ways to evaluate predictions that have distinct justifications. We report the results using in-sample R^2, out of sample R^2, Diebold and Mariano (2002) t-tests, and estimation with cross-validation.

Algorithm 2 GBRT Predictions

procedure GBRT($X_{i,t}, \pi_{i,t+1}$) ▷ Where t - time year, i - firm, $X_{i,t}$ - predictors at time t, $\pi_{i,t+1}$ - profitability at $t+1$ to predict
  for $1975 \leq T \leq 2014$ **do**
    Training sample are observations for all firms and $1964 \leq t \leq T-1$
    Fit the following GBRT model using training sample data
    $\pi_{i,t+1} = f(X_{i,t})$
    Get estimated model $\hat{f}_T$
    Use information at time T $X_{i,T}$ and fitted model $\hat{f}_T$ to predict profitability at $T+1$
    $\hat{\pi}_{i,T+1} = \hat{f}_T(X_{i,T})$
  end for
end procedure

4.1 In-sample R^2

An in-sample R^2 is perhaps the best known method to assess the ability of a model to account for the variation in the data. If a model is correctly specified, then maximum likelihood estimated parameters will be optimal. As argued by Inoue and Kilian (2005) it is then appropriate to evaluate those parameters within the sample.

Commonly we have less than full confidence that the structure of the model being estimated is the actual true data generating process. Any model may be reasonable, but it is almost certainly misspecified relative the true data generating process. Accordingly, as is standard in the machine learning literature (Efron and Hastie, 2016), we place much greater weight on the out-of-sample performance as discussed in the next section.

Even if we are not confident that the model is correctly specified, an in-sample R^2 is widely reported as a familiar diagnostic tool. In-sample explanatory power is given by

$$R^2 = 1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \bar{\pi}_{i,t+1})^2}. \tag{7}$$

where $\bar{\pi}_{i,t+1}$ is the sample average profitability. Because this is in-sample, it is bounded below by 0. The calculation steps are given in greater detail as Algorithm 3.

Algorithm 3 In-Sample R^2

procedure IS($X_{i,t}, \pi_{i,t+1}$)▷ Where t - time year, i - firm, $X_{i,t}$ - predictors at time t, $\pi_{i,t+1}$ - profitability at $t+1$ to predict

Use observations such that $1975 \leq t \leq 2014$ as the whole sample, ($1976 \leq t+1 \leq 2015$)

if evaluate FM **then**

for all firm-year observations in the whole sample **do**

for all time period t, run cross-sectional regression for all firms at time t

$\pi_{i,t+1} = \lambda_{0,t} + \lambda_{1,t} X_{i,t} + \varepsilon_{i,t}$

get estimated coefficients for each time period t $\{\hat{\lambda}_{0,t}, \hat{\lambda}_{1,t}\}$

compute the average coefficients $\{\hat{\lambda}_0, \hat{\lambda}_1\}$

end for

for all firm-year observations in the whole sample **do**

make prediction of profitability in the whole sample

$\hat{\pi}_{i,t+1} = \hat{\lambda}_0 + \hat{\lambda}_1 X_{i,t}$

calculate the in-sample R^2 in the whole sam

$R^2_{IS} = 1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \bar{\pi}_{i,t+1})^2}$ where $\bar{\pi}_{i,t+1}$ is the average of profitability $\pi_{i,t+1}$ for the whole sample

end for

end if

if evaluate GBRT **then**

for all firm-year observations in the whole sample **do**

Fit the following GBRT model using the whole sample

$\pi_{i,t+1} = f(X_{i,t})$

Get estimated model $\hat{f}$

end for

for all firm-year observations in the whole sample **do**

Use fitted model $\hat{f}$ to predict profitability in the whole sample

$\hat{\pi}_{i,t+1} = \hat{f}(X_{i,t})$

calculate the in-sample R^2 in the whole sam

$R^2_{IS} = 1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \bar{\pi}_{i,t+1})^2}$ where $\bar{\pi}_{i,t+1}$ is the average of profitability $\pi_{i,t+1}$ for the whole sample

end for

end if

end procedure

4.2 Out-of-sample R^2

Out-of-sample explanatory power is given by

$$R^2 = 1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \pi_{i,t})^2}. \tag{8}$$

If $\hat{\pi}_{i,t+1}$ is very far from $\pi_{i,t+1}$ then the numerator may be much larger than the denominator and so the term being subtracted may be larger than one. So the out-of-sample R^2 is not bounded below by zero. When that happens the model is clearly not doing a good job of explaining the data.

4.3 Diebold and Mariano (2002)

A popular and useful method of evaluating model performance is the Diebold and Mariano (2002) t-test, see also Diebold (2015). Following Gu et al. (2020), the test statistics DM is calculated as

$$\begin{aligned} DM &= \bar{d}_t^{1,2}/\hat{\sigma}(\bar{d}_t^{1,2}), \text{where} \\ d_{t+1}^{1,2} &= \frac{1}{n}\sum_{i=1}^{n}((e_{i,t+1}^2)^2 - (e_{i,t+1}^1)^2), \end{aligned} \tag{9}$$

$e_{i,t+1}^1, e_{i,t+1}^2$ is the prediction error for firm i profitability at time $t+1$ using each method, and $\bar{d}_t^{1,2}$ and $\hat{\sigma}(\bar{d}_t^{1,2})$ are the mean and Newey-West standard error of the time series $d_t^{1,2}$, respectively.

4.4 Cross-validation

Cross-validation is a standard machine learning method intended to reduce overfitting, see (Hastie et al., 2009; Efron and Hastie, 2016; Bates et al., 2021). With a panel of data there is again an issue of how to deal with the time dimension. We also want to ensure data comparability to the out-of-sample R^2 calculation.

Accordingly, firm-year observations from 1975 to 2014 are used as the data. The data is randomly partitioned into 10-folds using subsamples that are stratified by year. This ensures proper balance of years in each fold. No other factors were used for further stratification.

The model is estimated 10 times and then an average is reported. In the first round, the first fold is held out for validation and the remaining 9 folds are used for parameter estimation. In the second round, the second fold is held out for validation and the remaining 9 folds are used for parameter estimation. The process continues in this way until 10 estimates have been computed.

The reported out-of-sample R^2 for cross-validation is the average out-of-sample R^2 across all 10 estimations. The calculation steps are given in detail as Algorithm 4.

5 Evidence of predictability on average

This section provides overall evidence on the predictability of firm profits. Table 2 provides evidence on the efficacy of Fama-MacBeth and gradient boosting methods. Results are provided using the same factors used by Fama and French (2006) as well as the larger set of 149 factors as described in Section 2.

In column (1) of Table 2 prediction results for Fama and MacBeth (1973) regressions using Fama and French (2006) variables are reported. The model does extremely well with an in-sample R^2 of 0.65. In cross-validation and out-of-sample, the model does much less well than in-sample. This is consistent with the idea that over-fitting is taking place.

The cross-validation R^2 is much lower than the in-sample R^2, but it remains larger than the out-of-sample R^2. This strongly suggests that there remains a degree of overfitting. There are two likely reasons. In constructing the folds, we have only stratified on time. But the population of firms may have changed in other respects over time in ways that are not taken into account by the stratification. There is also the issue of possibly important macro factors that affect more recent time periods but not earlier time periods.

Column (2) again uses Fama-MacBeth regressions but extends the set of variables to include all 149 candidate factors. As might be expected, the model estimation collapses. There is substantial multicollinearity, and together with the overfitting found in column (1) it is easy to understand the inability of the model to handle the extra factors.

Algorithm 4 K-Fold Stratified Cross Validation

procedure CV($k, X_{i,t}, \pi_{i,t+1}$) ▷ Where t - time year, i - firm, $X_{i,t}$ - predictors at time t, $\pi_{i,t+1}$ - profitability at $t+1$ to predict, K-fold, stratified by time t

$K = 10$

Use observations such that $1975 \leq t \leq 2014$ as the full sample, $(1976 \leq t+1 \leq 2015)$

The sample is randomly split into K groups, $\{S_j\}_j^K$, stratified by year, meaning that each subsets contains roughly the same proportions of observation in each year. The K-Fold Stratified splitting is implemented using sklearn.model_selection.stratifiedkfold (scikit-learn 0.24.2).

for $1 \leq k \leq K$ **do**

the training set T_k is $\bigcup_{j=1, j\neq k}^{K} S_j$

the validation set V_k is $S_{j=k}$

if evaluate FM **then**

for all firm-year observations in the training set T_k **do**

for all time period t, run cross-sectional regression for all firms at time t

$\pi_{i,t+1} = \lambda_{0,t} + \lambda_{1,t} X_{i,t} + \varepsilon_{i,t}$

get estimated coefficients for each time period t $\{\hat{\lambda}_{0,t,k}, \hat{\lambda}_{1,t,k}\}$

compute the average coefficients $\{\hat{\lambda}_{0,k}, \hat{\lambda}_{1,k}\}$

end for

for all firm-year observations in the validation set V_k **do**

make prediction of profitability in the validation set V_k

$\hat{\pi}_{i,t+1} = \hat{\lambda}_{0,k} + \hat{\lambda}_{1,k} X_{i,t}$

calculate the out-of-sample R^2 at the validation set V_k

$R^2_{OOS,k} = 1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \pi_{i,t})^2}$ where $i, t \in V_k$

end for

end if

if evaluate GBRT **then**

for all firm-year observations in the training set T_k **do**

Fit the following GBRT model using training set T_k

$\pi_{i,t+1} = f(X_{i,t})$

Get estimated model $\hat{f}_k$

end for

for all firm-year observations in the validation set V_k **do**

Use fitted model $\hat{f}_k$ to predict profitability in the validation set V_k

$\hat{\pi}_{i,t+1} = \hat{f}_k(X_{i,T})$

calculate the out-of-sample R^2 at the validation set V_k

$R^2_{OOS,k} = 1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \pi_{i,t})^2}$ where $i, t \in V_k$

end for

end if

Compute the average out-of-sample R^2 for the model being evaluated

$R^2_{\text{K-Fold}} = \frac{1}{K} \sum_{k=1}^{K} R^2_{OOS,k}$

end for

end procedure

Table 2: Predicting profits overall

Profits are predicted for 1976 to 2015. Rolling estimation uses strictly prior data from 1964 to 2014 to make (pseudo) out-of-sample predictions. Cross validation and in-sample estimation uses data from 1975 to 2014 to make the predictions. The variable being predicted is $\frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. The amount of variation explained is denoted R^2. In-sample R^2 is calculated as $1-\frac{\sum(\hat{\pi}_{i,t+1}-\pi_{i,t+1})^2}{\sum(\pi_{i,t+1}-\bar{\pi}_{i,t+1})^2}$. For the cross validation we form 10 equally large groups stratified by year. 9 groups are used to estimate and then predict for the left out group. This is done 10 times. Then an average R^2 is computed for all 10 groups. Out-sample R^2 means that data from 1964 to 2014 is used to estimate the model on a rolling basis. The predictions are for 1976 to 2015. The models are fit using information from 1964 until time t to predict profits at time t+1. The out-of-sample R^2 is calculated as $1-\frac{\sum(\hat{\pi}_{i,t+1}-\pi_{i,t+1})^2}{\sum(\pi_{i,t+1}-\pi_{i,t})^2}$. Estimation using the Fama and MacBeth (1973) method is denoted FM. Estimation using the Friedman (2002) method is denoted GBRT. When the data used as explanatory variables follows Fama and French (2006) it is denoted FF06. When the data used as explanatory variables are all factors used in Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) it is denoted All. Data construction details are provided in the appendix. Every cell in this table has 121,401 observations.

	(1)	(2)	(3)	(4)
Estimation Method	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
In-Sample R^2	0.65	-18.67	0.68	0.72
10 fold CV R^2	0.06	-67.86	0.11	0.18
Out-of-Sample R^2	0.03	-2840.30	0.10	0.15

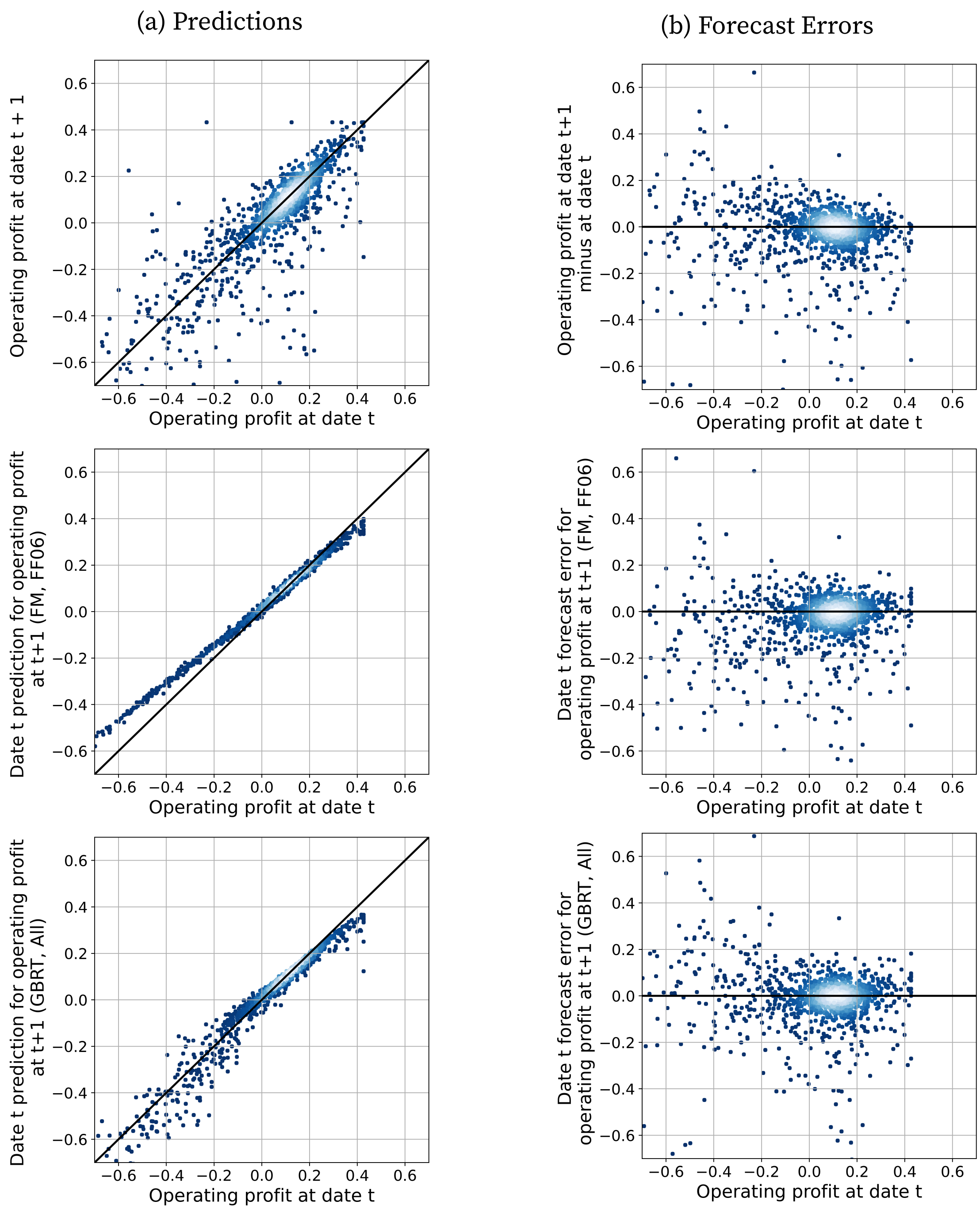
Column (3) uses gradient boosting with the FF06 factors. The model does predict better than column (1). There is again evidence of an important distinction between in-sample and out-of-sample performance. The cross-validation approach is again not sufficient to get fully remove the overfitting in-sample issue, but it is much closer than it was in column (1).

Column (4) uses gradient boosting but now with all of the 149 factors. Gradient boosting shows a major advantage over Fama-MacBeth. Instead of the model performance being destroyed as in column (2), there is now a significant improvement. This is true in-sample, using cross-validation, and also out-of-sample. Gradient boosting permits the use of extra factors beyond what can be effectively used in linear regressions.

The machine learning literature is very positive on the use of cross-validation to mitigate over-fitting. However, Table 2 still exhibits some over-fitting or perhaps omitted recent macro factors. There may also be an issue of data reuse, see Bates et al. (2021).

Figure 2: Actual and expected profits

This figure plots the relationship between current profit (t) and date t prediction for date t+1 profit, and the relationship between actual next period profit (t+1) and date t prediction for date t+1 profit. Date t predictions for date t+1 profit are predicted using Fama-Macbeth method with Fama and French (2006) variables (FM, FF06) and GBRT with all variables (GBRT, All). Date t forecast error is defined as realized profit at date t+1 minus date t predictions for date t+1 profit. We include only observations for $t = 2014$.



Further insight into the predictions and the forecast errors comes from Figure 2. In the top row on the left hand side of the page profit at date t is on the horizontal axis and date $t + 1$ is on the vertical axis. The profit data from one year to the next is very close to the $45°$ line. In the top row on the right hand side the change in profit from date t to $t + 1$ is given as the vertical axis while the actual data t profit is still the horizontal axis. The forecast errors appear to be clustered around zero. There is more dispersion for extremely negative values of date t operating profit.

The second row in Figure 2 uses the Fama-MacBeth method and FF06 factors to construct the forecasts. The forecasts and the forecast errors are plotted. The estimation method dramatically shrinks the variance of the predictions. As can be seen on the left-hand-side of the second row, the method produces forecasts that depart noticeably from the $45°$ line. This effect is particularly marked for firms with negative operating profits.

The third row of Figure 2 shows the corresponding plots using gradient boosting and All factors. Unlike Fama-MacBeth the estimates generally seem to lie along the $45°$ line, and there is more variation than in Fama-MacBeth estimates. Both models seem to have more difficulty making good predictions for firms with negative profits than they do for firms with positive current profits.

5.1 Are the prediction models significantly different?

In Table 2 sharply different R^2 values are obtained for alternative models. Are the differences large enough to be statistically significant at conventional levels of confidence? Table 3 provides statistical tests of the models from Table 2 against each other using Diebold and Mariano (2002) t-statistics. In each case a positive number indicates that the column model outperforms the row model. We consider 4 approaches to forecasting: 1) using the own date t value as prediction for date $t + 1$ ("Own lag"), 2) Fama-MacBeth with FF06 factors, 3) gradient boosting with FF06 factors, 4) gradient boosting with All factors.

Table 3 results provide strong support for the advantage of gradient boosting. Own lag is rejected in favor of all other models. gradient boosting with All factors is significantly better than any of the other models considered. This evidence reinforces what is reported

Table 3: Testing prediction models against each other out-of-sample

This table uses Diebold and Mariano (2002) t-statistics to compare rolling out-of-sample profit predictions from alternative models using data from 1964 to 2014 to predict profits for 1976 to 2015. In the estimation, the variable being predicted is $\frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. There are two versions of the explanatory variables. FF06 means that the explanatory variables are from Fama and French (2006). All means that the explanatory variable are from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. Own lag means that instead of estimation, the period t value of profit is used as a prediction of profits for t+1. Data construction details are provided in the appendix. FM means that the prediction model is estimated following Fama and MacBeth (1973). GBRT means that the prediction model is estimated following Friedman (2002). The model performance is evaluated using Diebold and Mariano (2002) test. Following Gu et al. (2020), the test statistics DM is calculated as $DM = \bar{d}_t^{1,2}/\hat{\sigma}(\bar{d}_t^{1,2})$, where $d_{t+1}^{1,2} = \frac{1}{n}\sum_{i=1}^{n}((e_{i,t+1}^2)^2 - (e_{i,t+1}^1)^2)$, $e_{i,t+1}^1, e_{i,t+1}^2$ is the prediction error for firm i profitability at time $t+1$ using each method, and $\bar{d}_t^{1,2}$ and $\hat{\sigma}(\bar{d}_t^{1,2})$ are the mean and Newey-West standard error of the time series $d_t^{1,2}$, respectively. The average differences $\bar{d}_t^{1,2}$ comparing column model with the row model are shown in the table. The test statistics DM are shown in the brackets. The test statistics are shown in the brackets. A positive number indicates that the column model outperforms the row model. Each cell is reports a cross-section average t-statistic for the predictions, so each cell has 40 observations. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	FM, FF06	GBRT, FF06	GBRT, All
Own Lag	3.60**	11.75***	17.75***
	(2.43)	(4.53)	(4.94)
FM, FF06		8.15***	14.15***
		(3.10)	(3.67)
GBRT, FF06			6.00***
			(4.19)

in Table 2.

5.2 Profit prediction horizon

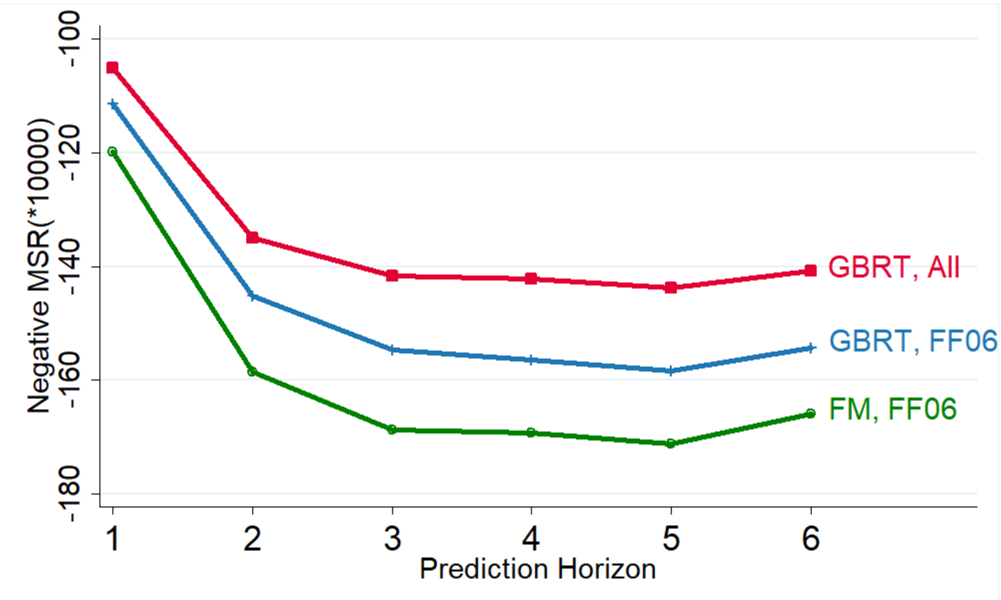
The tabulated results report results for one-year ahead profit predictions. Does predictability degrade rapidly as predictions are made about dates further in the future?

Figure 3 shows the results for date t predictions of profits for years $t+1$ to $t+6$ years. As expected, there is a sharp drop in predictability from year $t+1$ to $t+2$. Years $t+3$ to $t+5$ are essentially equally predictable. In year $t+6$ there is even a slight improvement in predictability. These horizon effects are similar across prediction methods and factors.

Consistent with Table 2 the gradient boosting method using all factors is consistently

Figure 3: Profit prediction horizon

This figure plots profit predictability at different prediction horizon. Future profits π_{t+j} with prediction horizon j is predicted using information at time t, X_t. The predictive models include: (1) Fama-Macbeth method with Fama and French (2006) variables (FM, FF06), (2) GBRT with Fama and French (2006) variables (GBRT, FF06), (3) GBRT with all variables (GBRT, All). The negative mean squared error (MSR) is calculated as $\frac{1}{N}\sum_{i,t}(\hat{\pi}_{i,t+j} - \pi_{i,t+j})^2$.



the best at predicting future profits at all horizons. The fact that longer horizon predictions beyond year 5 are not always worse than slightly shorter horizon predictions, is reminiscent of Fama and French (1988).

5.3 Hyperparameter tuning

The learning process of many machine learning models are controlled by parameters called hyperparameters. Unlike the ordinary parameters whose values are estimated directly by training the model on the data, hyperparameters are used to control the learning process. In the gradient boosting model, the learning rate γ ("learning_rate") determines the speed of convergence, the depth of the tree d ("max_depth") and the number of trees in the forest M ("n_estimators") control model complexity.

There is limited theoretical guidance on the choice of hyperparameters. The default values of the hyperparameters in standard libraries reflect satisfactory performance in many past applications. Our main estimation results use default values for the hyperparameters. In this section we ask, can the predictions be improved by refining the choices of the hyperparameters?

To find optimal hyperparameters we search over a hyperparameter space and then evaluate model performance. There are two common methods. It can be done using a hold-out validation sample or by using cross-validation. We try both, with quite similar results.

First, for hold-out validation the original data is divided into training, validation and testing samples. Models are estimated using alternative setting of the hyperparameters using the training data. In each case the performance of the hyperparameters is evaluated on the validation data. The best set of hyperparameters is selected. The final evaluation of the model is carried out using the testing data. To evaluate model performance we using out-of-sample R^2 computed on the testing data: $R^2_{OOS} = 1 - \frac{\sum_{it}(\pi_{i,t+1} - \hat{\pi}_{i,t+1})^2}{\sum_{it}(\pi_{i,t+1} - \pi_{i,t})^2}$.

Due to the panel structure of the data there is an issue of how best to reflect the passage of time. Following Gu et al. (2020) a recursive approach is used. For each year T, the training sample is $\{X_t, \pi_{t+1} : t \leq T-2\}$, the validation sample is $\{X_t, \pi_{t+1} : t = T-1\}$, the testing sample to predict π_{T+1} is $\{X_t : t = T\}$. The recursive sample splitting approach maintains the temporal ordering of the data.

Second, we also tune the hyperparameters using standard K-fold cross-validation. The sample is split into K smaller subsamples randomly. Due to the panel structure of the data, we stratify the samples by year so that the data does not accidentally overweight or underweight particular time periods. We do not stratify on factors other than time.

For each set of hyperparameters, we train the model using $K-1$ of the folds and then make prediction and compute the out-of-sample R^2 on the remaining fold of the training data. Each fold is used once as the held-out validation subsample and K out-of-sample R^2s are computed. The average over the K out-of-sample R^2 is the final performance measure. The set of hyperparameters that maximize the final performance measure is chosen as the optimal hyperparameters, and the final model is trained on the whole training subsample (all K-folds) using the best sum of squares residuals hyperparameters.

K-fold cross-validation method is computationally intensive. We tune the hyperparameters once at the beginning of the sample. We use information before 1975 $\{X_t, \pi_{t+1} : t \leq 1974\}$ as cross-validation sample, and maintain the tuned hyperparameters. van Binsbergen et al. (2020) also tune the hyperparameters once at the beginning of the sample, and maintain the tuned hyperparameter through out the paper.

The hyperparameters chosen by the K-fold cross-validation method is the following: learning rate is 0.1, max depth is 3, and number of boosting stages is 150. Apart from the number of boosting stages this is very much like the default hyperparameter values. Li and Rossi (2020) shows that different choices for the number of boosting stages does not significantly affect prediction performance.

Table 4 also shows that hyperparameter tuning does not affect out-of-sample R^2. The results suggest that different tuning method only generate slightly different out-of-sample performance. Given that we are interested in comparing the difference between Fama-MacBeth method and gradient boosting model, we opted to use the default parameters through the paper apart from Table 4. The prediction results are very similar when using alternative hyperparameters.

5.4 Which factors are most important?

The results above show that gradient boosting with All variables predicts profits better than does Fama-MacBeth. But the analysis included 149 factors. Which factors matter most? How do those compare to the factors used by Fama and French (2006)?

In the Fama-MacBeth model, the average slope coefficient and its variation, reflects the importance of each factor. But gradient boosting is a non-parametric model without a specific coefficient that plays a similar role to the average slope coefficient.

To evaluate the impact of individual factors we use feature importance, and a permutation test of feature importance. These two methods have distinct statistical foundations.

First consider feature importance. Following Hastie et al. (2009) we use an impurity-based measure also known as the Gini importance. It is computed as the total criterion reduction (Gini Gain) brought by each feature. The Gini index is similar to an empirical

Table 4: Does hyperparameter tuning make much difference?

This table provides out-of-sample performance for different hyperparameter tuning methods. The variable being predicted is $\frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. Estimation using the Friedman (2002) method is denoted GBRT. The amount of variation explained is denoted R^2. When the data used as explanatory variables are all factors used in Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) it is denoted All. Hyperparameters are set in three ways: (1) defaults: learning rate is 0.1, max depth is 3, and number of boosting stages is 100, (2) recursive evaluation: use $\{X_t, \pi_{t+1} : t <= T-2\}$ as the training sample, $\{X_t, \pi_{t+1} : t = T-1\}$ as the validation sample, and $\{X_t : t = T\}$ as the testing sample to predict π_{T+1}, (3) 10-fold stratified cross validation stratified by year: sample before 1975 $\{X_t, \pi_{t+1}\ t <= 1974\}$ is used as cross-validation sample, and maintain the tuned hyperparameters for the rest. For the methods in (2) and (3), the hyperparameters are tuned over the following ranges: learning over [0.01, 0.1, 0.2], max depth over [1,2,3,4], and n estimators over [50, 100, 150]. For the column (3) 10-Fold cross validation, the hyperparameters chosen by the K-fold cross-validation method is: learning rate is 0.1, max depth is 3, and number of boosting stages is 150. Out-of-sample R^2 values calculated as $1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \pi_{i,t})^2}$. For out-of-sample R^2 calculations, every cell has 121,401 observations. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

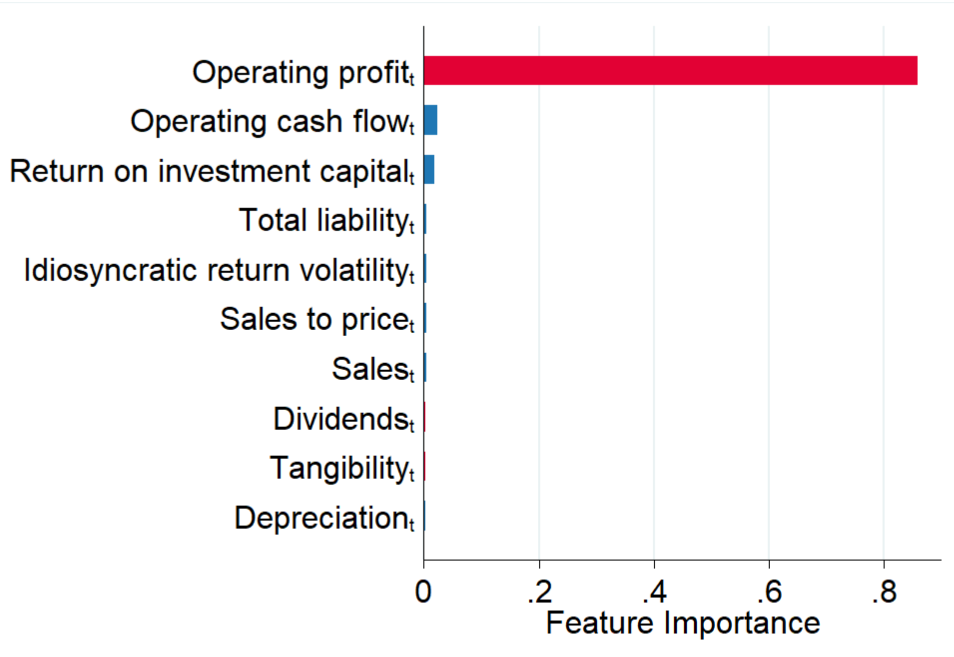
	(1)	(2)	(3)
Estimation Method	GBRT	GBRT	GBRT
Data	All	All	All
Tuning	Default	Recursive	Cross Validation
Out-of-Sample R^2	0.15	0.13	0.15

variance under the node. Gini Gain is the variance reduction after introduce x_m as splitting variable. The feature importance of regression tree for feature V is the sum of Gini Gain of the variable across all nodes within the tree. We take the weighted-average over ensemble of trees. Then we apply the feature importance at the ensemble or 'forest' level. Feature importance is calculated for each variable in the gradient boosting model. Figure 4 shows the ratio of each factor's feature importance overall the sum of total feature importance. The values for the top 10 features are shown. If a feature is included as a factor by Fama and French (2006) it is shown using a red bar. If not, it is shown using a blue bar.

By far the most important single factor for predicting period $t+1$ profit is period t profit. It constitutes 84% of the total feature importance from all factors. This is the one top factor that was also used by Fama and French (2006). Despite the importance of period t profits, recall the result in Table 3. The period t profit is not a sufficient statistic for predicting period $t+1$ profit. That model is rejected relative to all the other models considered.

Figure 4: Measuring feature importance

This figure plots average feature importance of variables of the GBRT model. The feature importance is calculated as the improvement in accuracy brought by each predictor. Improvement in accuracy is defined as the decrease in mean squared errors. The x-axis is the average feature importance. Feature importance is standardized such that the total feature importance of all predictors are sum up to one. Red color indicates that the feature belongs to FF06 variables.



The overlap between the FF06 factors and the top 10 feature importance factors in Figure 4 is surprisingly limited. The remaining factors are all fairly reasonable, but have much smaller impacts. In order we have operating cash flow, return on investment capital, total liability, and so forth.

Next consider the permutation test of feature importance. It is also known as "Mean Decrease Accuracy", see Louppe et al. (2013). This is another common method to examine the impact of individual factors.

The permutation importance of a predictor is defined as follows. First a model is trained on the training sample. An evaluation metric is calculated as baseline metric. We use R^2 as the evaluation metric. Second, using testing data, a predictor column is randomly permuted. Third, the trained model to make predictions on the permuted testing sample, and use the same evaluating metric to calculate the performance of the model on the permuted testing sample. Finally, the difference between the baseline metric and evaluating metric on the permuted testing sample is defined as the permutation importance of the predictor.

The random permutation for each predictor can be done N times. To implement the permutation tests, we use the Python software library https://eli5.readthedocs.io/en/latest/ with the method: PermutationImportance.

Conceptually, impurity-based feature importance and permutation importance are rather different from each other. First, the calculations of impurity-based importance are based solely on the training sample. Therefore, impurity-based feature importance does not directly reflect the predictive power of the model on the testing sample. Second, impurity-based measure is dependent on the cardinality of the predictor (Louppe et al. (2013)) and is biased toward high cardinality predictors (Cerda and Varoquaux (2020)). In contrast, permutation importance focuses on the predictions.

Despite the conceptual differences, the evidence in Figure 5 is reasonably similar to that in Figure 4. Once again period t profit is by far the dominant factor. The ordering of the subsequent factors is not the same as in Figure 4, although a number of the same factor remain in the top 10 list.

6 Evidence of differences across firms

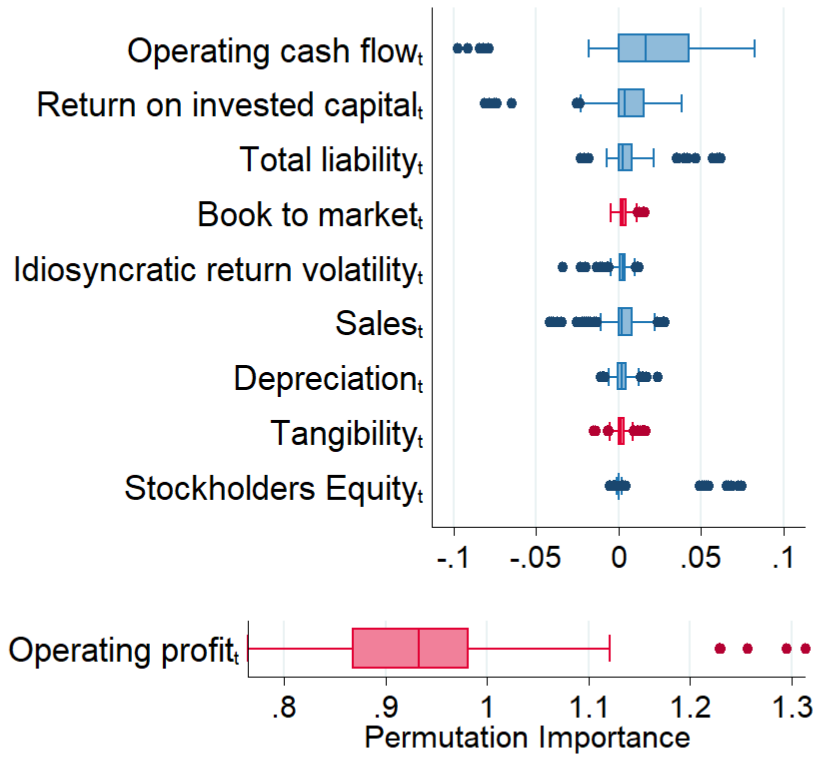
The results so far establish that average firm profits are predictable. But there is considerable heterogeneity among firms. Accordingly, in the section we focus on that heterogeneity.

6.1 Predicted profit differences

Among the most frequently studied difference among firms is the impact of firm size. In Table 5 firms are sorted into quintiles by total assets on a date t. Within each quintile average profits for date t, and date $t + 1$ are reported along with the predicted date $t + 1$ profits using the FM, FF06 model and using the gradient boosting, All model.

Figure 5: Permutations tests of feature importance

This figure plots the permutation importance distribution of variables of the GBRT model. The permutation importance of a predictor is calculated as the average decrease in accuracy when the predictor value is randomly shuffled. Each predictor is randomly shuffled 30 times. The x-axis is the permutation importance. The distribution is plot as box plots. Permutation value is standardized such that the total permutation of all predictors are sum up to one. Red color indicates that the feature belongs to FF06 variables.



Within each quintile both date (t) and date ($t+1$) profits are monotonic increasing as firm size increases until we reach quintile 4. Quintiles 4 and 5 have essentially the same profitability relative to assets. The differences between the top and the bottom quintiles are statistically significant.

Expected profits show similar patterns for expected profits across estimation methods and factors. However the FM, FF06 model slightly over predicts average profits in quintiles 1, 2 and 3. The gradient boosting, All model does a better job of matching average profits over these three quintiles. Both models do a good job in the top size quintile 5. So the advantage of gradient boosting relative to FM seems to be associated with doing

Table 5: Are large firms more profitable?

This table reports profitability for firms sorted into quintiles annually based on total assets at date t. "Smallest" indicates that smallest assets quintile, "Largest" indicates the largest assets quintile. "L-S" reports the average difference between the smallest firm quintile and the largest. The t-stats are for tests that the "L-S" values are different from zero. The data and estimation methods are the same as in Table 2. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Quintile	(1)	(2)	(3)	(4)	(5)		
	Smallest				Largest	L - S	t-stat
Total assets$_t$	21.56	79.56	229.82	725.17	7,483.72	7,462.17***	(89.89)
Observations	24,299	24,278	24,282	24,278	24,265		
Profit$_t$	-0.00	0.09	0.13	0.14	0.14	0.14***	(77.62)
Profit$_{t+1}$	-0.01	0.07	0.12	0.14	0.14	0.15***	(77.77)
E_tProfit$_{t+1}$ (FM, FF06)	0.02	0.09	0.13	0.14	0.14	0.12***	(81.84)
E_tProfit$_{t+1}$ (GBRT, All)	-0.00	0.08	0.12	0.13	0.14	0.14***	(89.30)

a better job of predicting average profits for smaller firms. For the largest firms there is less of an advantage. Overall, the data says that larger firms are more profitable and the models capture this fact.

Table 6 examines firms that exit. It tabulates profits and expected profits for firms that exit from the sample before the final year of our data (2015). The population average profit (operating profit/ta) across all firm and all time periods is 0.09.

Compustat reports several reasons for a firm to exit. Let t be the final year that the firm exists in our data. In each case that a reason is reported, we provide the average profitability for the previous year $t-1$, now t and the expected profits according to our preferred model (GBRT, All).

There are noteworthy differences. Firms that are bankrupt or liquidated have much lower profits and would be expected to have low (or negative) profits had they not exited. That is also true of firms in a 'reverse acquisition'. A reverse acquisition happens when a private firm buys a public firm in order to go public while avoiding an IPO.

Firms in an ordinary merger, acquisition or an LBO have positive profits. Only the LBO firms have higher profits than an average publicly traded firm. These connections

Table 6: Profits and reasons for exit

This table reports profitability differences for firms that exit from the data before the final year of the data, 2015. Year t is the last year that the firm is in the data. So $E_t\pi_{t+1}$ is the profit that prediction as of year t for the firm, if it had not exited. It is based on the final available data prior to the exit. The Compustat deletion code reported as footnote 35 gives a measure of why a firm exits. Population average is defined as the mean value of operating profitability of all firm-year observations. Previous means the year before the final year that the firm is in the data. Now means the last year that the firm is in the data. Expected is a forecast based on GBRT estimation using Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) variables together and estimation by Friedman (2002) method. Average is the population average profitability. *, **, and *** denote the value is statistically significantly different from the population average at the 10%, 5%, and 1% level, respectively.

	population average $\pi = 0.09$			
	π_{t-1}	π_t	$E_t\pi_{t+1}$	Obser
Bankruptcy	0.01***	-0.04***	0.00***	463
Liquidation	0.01***	-0.06***	-0.02***	317
Reverse acquisition	-0.15***	-0.19***	-0.16***	132
Acquisition or merger	0.08***	0.07***	0.07***	7174
Leveraged buyout	0.15***	0.15***	0.14***	73
Now a private company	0.09	0.07**	0.08*	336

between expected firm profit and the reason for firm exit seem reasonable.

6.2 Predictability differences

Table 7 considers the connections between a number of firm attributes and the predictability of profits using minus the sum of squared residuals as our definition of predictability. Firms are sorted into quintiles based on predictability and then a range of firm attributes are averaged within each quintile. Finally we test the hypothesis that the mean values of the attribute are the same in the highest predicability quintile and the lowest predictability quintile. This is done for GBRT, All, for FM, FF06, and then we also consider the difference between these estimates.

High profit firms are much more predictable both using FM, FF06 and using GBRT, All. Firms with high R&D to sales, market-to-book and cash flow volatility are all less predictable. The results are generally stronger under gradient boosting, All than under the FM, FF06 model. However the basic predictability patterns are the same across methods.

Table 7: What type of firms have more predictable profits?

This table reports several firm attributes for firms sorted into quintiles annually based on profits predictability. For each firm year, the predictability is calculated as negative squared error (SR), $-(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2$. In columns (1) - (3), the prediction model is estimated following Friedman (2002), with the explanatory variables from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. In columns (4)-(6), the prediction model is estimated following Fama and MacBeth (1973), with explanatory variables from Fama and French (2006). In columns (7)-(9), firms sorted into quintiles annually, based on the predictability difference between GBRT, All model and FM, FF06 model. $H - L$ is the difference between the highest and lowest decile value. The t-stats in the brackets are for tests that the $H - L$ values are different from zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	GBRT, All			FM, FF06			GBRT, All - FM, FF06		
	Lowest	Highest	H - L	Lowest	Highest	H - L	Lowest	Highest	H - L
Negative $\text{SR}_t + 1$	-483.28	-0.29	482.99***	-557.75	-0.31	557.44***	-87.95	160.34	248.30***
			(62.26)			(60.11)			(63.99)
$\text{Profit}_t + 1$	-0.04	0.13	0.16***	-0.05	0.13	0.18***	0.09	-0.04	-0.12***
			(75.94)			(87.76)			(-54.30)
Profit_t	-0.00	0.13	0.13***	-0.01	0.13	0.14***	0.05	0.03	-0.03***
			(70.90)			(73.24)			(-11.92)
Total Assets_t	490.00	2976.01	2486.00***	508.02	2829.15	2321.13***	694.84	805.16	110.32***
			(41.57)			(40.13)			(2.91)
R&D to sales_t	0.22	0.10	-0.76***	0.83	0.06	-0.77***	0.47	0.62	0.15***
			(-26.46)			(-26.62)			(4.72)
Market to Book_t	2.23	1.45	-0.78***	2.18	1.48	-0.70***	2.08	2.04	-0.04**
			(-49.99)			(-45.97)			(-2.18)
Cash Flow Volatility_t	13.48	1.86	-11.62***	13.70	1.76	-11.94***	7.99	11.83	3.84***
			(-18.10)			(-17.25)			(5.00)
Observations	24281	24280	48561	24281	24280	48561	24281	24280	48561

6.3 Predictability during recessions

It is well known that firms commonly make less profits during recessions. But does the predictability of profit also change? We examine 3 methods of prediction in Table 8: GBRT, All; FM, FF06; and date t profit as a prediction of date $t+1$ profits. In all cases we use minus the squared difference between actual π_{t+1} and $E_t\pi_{t+1}$ as the dependent variable.

Table 8 Panel A, examines what happened during the financial crisis of Dec 2007 to June 2009. During the crisis period there is a sharp drop in profit predictability according to all models. The effect is strongest in the FM, FF06 model. In all models the drop in predictability is much less acute for investment grade firms. Suppose that in a financial crisis investors are looking for predictable profits – a version of a flight to quality. Then we should see them selling average firms and buying investment grade firms. Since supply and demand for shares must still be equal, the relative prices would adjust with average firms falling relative to investment grade firms.

Panel B extends the tests to all NBER recessions that took place during our sample period. The results are very similar to those in Panel A. In this respect the financial crisis looks very similar to any other recession. Overall we see that on average firm profits are less predictable during recessions. This effect is large for an average firm, but it is much smaller for investment grade firms.

7 Predictable profits and the flow of financial resources

In this section we ask whether financial resources tend to flow to more profitable firms. Using our estimates we can distinguish the impact of actual current profits and expected future profits. While the supply and demand for existing shares must be equal, firms can issue or repurchase shares and debt securities thereby changing the volume of securities that the firm has outstanding.

According to the model in Frank and Sanati (forthcoming) the distinction between actual profit and expected profit is critical for the flow of finance. In their model a firm that

Table 8: Predicting profits in recessions

This table reports OLS regressions with 'Predict' as the dependent variable. It is a measure of how predictable profits are and it is measured as the negative squared difference between predicted next period profit and actual next period profit. The predicted profits are estimated as in Table 2. In columns (4), the dependent variable is the difference in profit predictability between the GBRT, All model and FF, FF06 model. InvGrade is an indicator variables that equals to one if the firm in a given year has S&P Domestic Long Term Issuer Credit Rating better or equal to BBB. Recession is an indicator variables that equals to one if the data date ended in the NBER recession period. Crisis is an indicator variable that equals to one if the data date ended in the NBER recession from Dec 2007 to June 2009 which is often called the Great Recession. Standard errors are clustered at the firm level, and t-statistics are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every column in this table has 23,262 observations, which are the observations with available long term bond rating.

Panel A	Great Recession Dec 2007 - June 2009			
InvGrade*Crisis	17.93**	17.89**	24.07**	0.05
	(2.20)	(2.05)	(2.34)	(0.03)
InvGrade	22.32***	23.19***	26.12***	-0.87
	(6.69)	(6.52)	(6.82)	(-1.43)
Crisis	-19.44**	-21.97***	-29.12***	2.53
	(-2.48)	(-2.61)	(-2.93)	(1.54)
Observations	23185	23185	23185	23185
Method	GBRT, All	FM, FF06	Profit_t	(GBRT, All) - (FM, FF06)
Adjusted R^2	0.003	0.003	0.004	0.000

Panel B	NBER Recessions			
	Predict_{t+1}	Predict_{t+1}	Predict_{t+1}	Predict_{t+1} difference
InvGrade*Recession	19.06*	19.41*	23.43**	-0.36
	(1.88)	(1.84)	(2.09)	(-0.23)
InvGrade	21.74***	22.61***	25.61***	-0.87
	(6.54)	(6.36)	(6.66)	(-1.40)
Recession	-19.40*	-21.62**	-26.19**	2.21
	(-1.95)	(-2.09)	(-2.39)	(1.49)
Observations	23185	23185	23185	23185
Method	GBRT, All	FM, FF06	Profit_t	(GBRT, All) - (FM, FF06)
Adjusted R^2	0.003	0.003	0.004	0.000

has a positive productivity shock would like to invest more to take advantage. If unconstrained, the firm would issue debt to pay for the capital and exploit the tax advantage. However due to financing constraints the firm does not have adequate capital. So when

Table 9: Profits and financing flows

This table reports the effect of profitability and expected profitability on debt and and equity issuance and repurchasing. GBRT is used to forecast next period operating profit, π_{t+1}, using prior information available at time t including the variables from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. A constant term, year fixed effect, firm fixed effect, and firm control variables are included (Frank and Goyal, 2009). Standard errors are clustered at the firm level, and t-statistics are in parentheses. The dependent variable is always based on information that was available prior to the start of the time period. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every column in this table has $119,955$ observations.

	Debt$_t$			Equity$_t$			Assets Growth
Dep var	Net Iss	Gross Iss	Repur	Net Iss	Gross Iss	Repur	$\frac{\Delta_{t,t+1}AT}{AT_t}$
Profit$_t$	0.06***	-0.02*	-0.08***	-0.25***	-0.25***	-0.00	-0.10*
	(6.72)	(-1.95)	(-8.47)	(-12.95)	(-12.92)	(-0.63)	(-1.85)
E_tProfit$_{t+1}$	-0.02**	0.03**	0.06***	0.07***	0.10***	0.04***	0.49***
	(-2.23)	(1.99)	(4.70)	(2.85)	(4.25)	(8.98)	(7.00)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.07	0.39	0.40	0.46	0.46	0.23	0.23

profits are expected the firm issues equity and repurchases debt in order to buy more capital. Once it has the extra capital, together with the positive productivity shock the firm generates more actual profits. When the firm has actual profits and more capital as collateral, it issues debt to take average of the tax benefit and it repurchase the expensive equity.

Table 9 produces a pattern of coefficients on expected and actual profits are similar to the predictions of Frank and Sanati (forthcoming). A firm with high expected profit tends to reduce debt and raise funds by issuing equity. As time passes and the expected profits turns (on average) into actual profits the firm's financing constraints are relaxed. Now the firm uses the profits to repurchase equity. Because the constraints have been relaxed the firm issues debt. That permits the firm to take advantage of the tax benefit of the debt, as in Frank and Sanati (forthcoming).

The final column of Table 9 shows that expected future profit has a stronger effect on asset growth than does actual current profit. This is consistent with the forward-looking positive NPV investment theory that is commonly taught.

8 Predictable profits and the cross-section of stock returns

It is well known that measures of firm profit are connected to stock returns (Fama and French, 2006; Novy-Marx, 2013; Fama and French, 2015). Because gradient boosting with All factors has reasonable predictive power, we examine it's potential use for this purpose. We follow Fama and French (2006) and Aharoni et al. (2013) to test how well expected profit explains cross-sectional stock returns.

Using the notation from Fama and French (2006) and Aharoni et al. (2013), the firm's valuation equation is,

$$\frac{M_t}{B_t} = \sum_{\tau=1}^{\infty} \frac{E(Y_{t+\tau}/B_t) - E(\Delta B_{t+\tau}/B_T)}{(1+r)^{\tau}}$$

where M_t is the market value of equity, B_t is the book value of equity, Y_t is profit, r is the average expected stock return.

Controlling for $\frac{M_t}{B_t}$ and expected growth in book equity, more profitable firms have higher expected returns. Therefore, adding expected profit and expected asset growth could help explain the cross-sectional stock returns. So the key idea for this purpose is that a better expected profit measure should explain better the cross-sectional stock returns. To examine the efficacy of the expected profit measures we follow the procedure developed and used in Fama and French (2006) and Aharoni et al. (2013).

Portfolios are constructed based on predicted return according to models using the distinct expected profit measures. The portfolios are sorted into quintiles based on the predicted returns. Then we examine the realized return differences between portfolios of stocks with high predicted returns vs portfolios of stocks with low predicted returns. If a model is better at explaining cross-sectional returns, then there will be a larger realized return differences between the high and the low portfolios.

Predicted return are estimated monthly using Fama-MacBeth regressions for the July

1975 to December 2015 sample period.

$$r_{i,t+1,k} = \theta_0 + \theta_1 \ln bkmkt_{i,t} + \theta_2 \ln mv_{i,t+1} + \theta_3 E_t \pi_{t+1} + \theta_1 \ln E_t \Delta_{t+1} AT/AT_t$$

To test the explanatory power of different expected profit measures in the valuation equation framework, we add expected profit and expected asset growth to predict return when allocating stocks. To be consistent, we measure expected asset growth estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors. Column (1) uses size and market-to-book only as a benchmark.

Column (2) uses profit on date t as expected profit measure ($E_t \pi_{t+1} \equiv \pi_t$). Column (3) measures expected profit estimated using Fama-MacBeth regressions with FF06 factors. Column (4) uses gradient boosting with All factors to measure expected profits.

Column (2) using date t profit as expected profit, finds that the monthly return spread between high and low portfolio is 0.82%. This is larger than the return spread in column (1). The difference is 0.12% per month.

Using Fama MacBeth to measure expected profit in column (3), the return spread increase from 0.82% per month (in column 2), to 0.89% per month. Column (4) uses gradient boosting with All factors produces a similar spread to column (3). To some degree this apparently this reflects the use of quintiles. In untabulated results we did the same analysis but using 25 portfolios instead of quintiles. In that cased gradient boosting with All factors generates a larger spread than does Fama-MacBeth with FF06 factors.[1]

The results for the abnormal return spread seem reasonable. Existing literature in profitability premium (Novy-Marx, 2013; Wang and Yu, 2013) has documented that the profitability premium exists primarily among firms with low book-to-market ratios and firms with high information uncertainty. Table 7 shows that the improvement of gradient boosting-based expected profits are concentrated among firms with low book-to-market

[1] The results are similar when we followed Novy-Marx (2013) and calculated value-weighted portfolio excess returns and the α from a 3-factor model that is sorted on expected profits, see Table C8. The results are not totally due to denominator effects (e.g., Ball et al. (2015)). When we define profitability as the operating profit scaled by lagged total assets, the same pattern of results is found from Table C9. Moreover, in Table C12 we find that the improvement of GBRT model is stronger for income deflated by the book value of equity.

Table 10: Profits and the cross section of expected stock returns

This table presents the monthly value-weighted average realized returns and spreads of portfolios formed on predicted returns. The predicted return are estimated monthly using Fama and MacBeth (1973) regressions for July 1976 to June 2016 sample period.

$$r_{i,t+1,k} = \theta_0 + \theta_1 \ln bkmkt_{i,t} + \theta_2 \ln mv_{i,t+1} + \theta_3 E_t \pi_{t+1} + \theta_1 \ln E_t \Delta_{t+1} AT/AT_t$$

$r_{i,t+1,k}$ is the return on stock i in the kth month of the 12 months from the July of calendar year $t+1$ to the June of calendar year $t+2$. i also denotes stock i in all the independent variables. Book-to-market, $\ln bkmkt_{i,t}$, is the logarithm of the book value of equity at the end of the fiscal year that ends in calendar year t divided by the market value of equity at the end of calendar year t. Size, $\ln mv_{i,t+1}$, is the logarithm of the market value of equity at the end of June of calendar year $t{+}1$. Expected profitability, $E_t \pi_{t+1}$, is the expected value of profit in the fiscal year ending in calendar year t+1. Expected asset growth, $E_t \Delta_{t+1} AT/AT_t$, is the expected growth of total assets in the fiscal year ending in calendar year t+1. In column (1) independent variables used to predict return are size and book-to-market only, which is the benchmark portfolio. In columns (2)-(4), we add expected asset growth and expected profitability. To be consistent, we measure expected asset growth estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors. In columns (2), we use profit now as expected profit measure ($E_t \pi_{t+1} \equiv \pi_t$). In columns (3), we measure expected profit estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors, In columns (4), we measure using GBRT and all predictors. The expected profits are estimated similarly as in Table 2. Predicted return from July of year t to June of year t+1, the fitted value from the Fama and MacBeth (1973) regression equation, are the product of average regressions slopes and explanatory variables at the end of June of year t. Stocks are sorted into quintiles according to their predicted return. Value-weighted average return is calculated for each group. We report the average realized returns of the portfolio with the lowest predicted return (Low) and the portfolio with the highest predicted return (High). The returns are in percentage points, so for example 2, means the monthly return is 2%. We calculate the spread between the highest and lowest predicted return portfolios (High - Low). We also compute the average difference and t-test statistics (in the brackets) between the “High - Low" spread in each column and the benchmark "High - Low" spread in column (1). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every portfolio has 480 monthly observations.

	(1)	(2)	(3)	(4)
Portfolio	size+value	Profit$_t$	E_tProfit$_{t+1}$	E_tProfit$_{t+1}$
Low	0.55	0.52	0.53	0.53
High	1.26	1.34	1.42	1.42
High - Low	0.70	0.82	0.89	0.89
Aver diff		0.12	0.19	0.19
t-statistic		1.68	2.69	2.90
Method			FM,FF06	GBRT,All

ratio and high cash-flow uncertainty. So it seems reasonable that a gradient boosting based measure of expected profits could improve the empirical performance of the profitability premium among firms that are particularly sensitive to that factor.

9 Predictable profits prediction mistakes?

An alternative approach to consider profit predictions is to connect the predictions to rational expectations. Nagel (2021) has suggested that machine learning predictions provide a reasonable as-if model of the expectations of actual investors. On the other hand Bordalo et al. (2021), find that people have 'diagnostic expectations' and so they overreact to shocks. If Nagel (2021) is right, then the algorithms ought to generate the same coefficient patterns found among stock analysts by Bordalo et al. (2021). Is that actually true of the data? This section examines this hypothesis.

For comparability we adopt the framework used by Bordalo et al. (2021), see also Afrouzi et al. (2020). Specify,

$$\mathbb{E}_t^{\theta}(\pi_{t+1}) = \mathbb{E}_t(\pi_{t+1}) + \theta[\mathbb{E}_t(\pi_{t+1}) - \mathbb{E}_{t-1}(\pi_{t+1})] \tag{10}$$

where π_{t+1} is profits at date $t+1$, $\mathbb{E}_t(\cdot)$ is the rational expectations at date t. They call $\theta \geq 0$ a diagnosticity parameter. They say that $\theta = 0$ means rational expectations, while $\theta > 0$ means that "agents overreact to news, becoming too optimistic after good news and too pessimistic after bad news".

Bordalo et al. (2021) run linear regressions in which the dependent variable is the error at date $t+1$ of stock analyst predictions from IBES. The independent variables are date t values of the forecast, profit, investment, or debt issuance. In each of these regressions firm and year fixed effects are included. According to Bordalo et al. (2021) under rational expectations the slope coefficients should be zero. Under diagnostic expectations the slope coefficients should be negative. That is because of beliefs that overreact both to good and to bad news. Bordalo et al. (2021) use linear regressions with fixed effects to carry out this test. We follow Bordalo et al. (2021) in Table 11. We also add a column (1)

Table 11: Are prediction errors linearly predictable?

This follows Table 1 in Bordalo et al. (2021), regressing prediction error at $t+1$ on the information at time t. Prediction mistake at $t+1$ is Prediction Error$_{t+1}$ $:= p_{t+1} - E_t p_{t+1}$, where E_t is the model prediction based on date t data. Panel A uses Fama and MacBeth (1973) estimation and FF06 variables as the prediction model. Panel B uses GBRT and All variables. Four different information available at time t are used. Prediction Error$_t$ $:= p_t - E_{t-1}p_t$ is prediction error at time t. Profit$_t$ $= \frac{\text{operating profit}(t)}{\text{total assets}(t)}$ is profitability at time t. Investment$_t$ $= \frac{\text{capital expenditure}(t)}{\text{total assets}(t)}$ is investment rate at time t. Debt issuance$_t$ is net debt issuance at time t. Column (1) has fewer observations due to unavailable forecasts when firm enters the sample at time t.

	(1)	(2)	(3)	(4)	(5)
Panel A	Prediction error$_{t+1}$				
Prediction Error$_t$	-0.24***				
	(-24.63)				
E$_t$Profit$_{t+1}$		-0.37***			
		(-30.10)			
Profit$_t$			-0.29***		
			(-30.10)		
Investment$_t$				-0.02**	
				(-2.03)	
Debt issuance$_t$					-0.02**
					(-2.19)
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Method	FM,FF06	FM,FF06	FM,FF06	FM,FF06	FM,FF06
Observations	106290	119955	119955	119955	119955
Adjusted R^2	0.16	0.21	0.21	0.13	0.13
Panel B	Prediction error$_{t+1}$				
Prediction Error$_t$	-0.13***				
	(-14.43)				
E$_t$ Profit$_{t+1}$		-0.31***			
		(-28.76)			
Profit$_t$			-0.21***		
			(-23.41)		
Investment$_t$				-0.02***	
				(-2.85)	
Debt issuance$_t$					0.05***
					(6.11)
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Method	GBRT,All	GBRT,All	GBRT,All	GBRT,All	GBRT,All
Observations	106290	119955	119955	119955	119955
Adjusted R^2	0.11	0.15	0.14	0.10	0.10

that uses the prediction error at date t as an alternative regressor. Columns (2) - (5) use the specifications from Bordalo et al. (2021) but using the algorithms to generate the predictions being evaluated.

Table 11 Panel A provides results for profit predictions generated using Fama-MacBeth, FF06. Panel B provides results for profit predictions using gradient boosting, All. Under this testing framework the predictability of the prediction errors are remarkably similar.

In Panel A, all five columns produce statistically significant linear predictability of the prediction errors. In all cases the coefficient is negative and statistically significant. This is the same pattern of results reported by Bordalo et al. (2021) for stock analysts. Rational expectations is sharply rejected in favor of "overreaction".

Panel B carries out the same regressions as Panel A, but using the predictions generated by gradient boosting. There is a marginal reduction in the explanatory power of the regressions compared to Panel A. However, the results in columns (1) to (4) are again all negative and all statistically significant. In column (5) debt issuance is used as a factor and the sign reverses. It is again statistically significant. There is no obvious reason for this sign reversal, and it is likely simply a reflection of sampling variation. The key result of Panel B is that as in Panel A, rational expectations is rejected, and the bulk of the evidence is similar to Bordalo et al. (2021).

To dig more deeply into the nature of the linear predictability of the prediction errors, Table 12 redoes the prediction models, but introduces the date t prediction mistake as a factor when predicting date $t+1$ profits. For comparison Panel A provides the baseline results.

Notice that the Table 12 Panel A is not quite the same as Table 2. In Table 2 the prediction period is 1975 to 2015. In Table 12 the prediction period is 1991 to 2015. In Table 12 the prediction errors are only available after 1975. In order to include as many predictions errors as possible, in Panel B we make predictions from 1991 to 2015. To make Panel A consistent with Panel B, we use the same prediction period, which is why it differs from Table 2.

In Table 12 columns (1) and (2) use Fama-MacBeth estimation and columns (3) and (4) use gradient boosting. The results are very sharp. In columns (1) and (2) the out-of-

Table 12: Current prediction errors as a factor

Profits are predicted for 1991 to 2015. Rolling estimation uses strictly prior data from 1974 to 2014 to make (pseudo) out-of-sample predictions. Cross validation and in-sample estimation uses data from 1990 to 2014 to make the predictions. The variable being predicted is $\frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. The amount of variation explained is denoted R^2. In-sample R^2 is calculated as $1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \bar{\pi}_{i,t+1})^2}$. For the cross validation we form 10 equally large groups stratified by year. 9 groups are used to estimate and then predict for the left out group. This is done 10 times. Then an average R^2 is computed for all 10 groups. Out-sample R^2 means that data from 1974 to 2014 is used to estimate the model on a rolling basis. The predictions are for 1991 to 2015. The models are fit using information from 1974 until time t to predict profits at time t+1. The out-of-sample R^2 is calculated as $1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \pi_{i,t})^2}$. Estimation using the Fama and MacBeth (1973) method is denoted FM. Estimation using the Friedman (2002) method is denoted GBRT. When the data used as explanatory variables follows Fama and French (2006) it is denoted FF06. When the data used as explanatory variables are all factors used in Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) it is denoted All. In Panel B, we add the current prediction error defined in Table 11 as an additional predictor. Data construction details are provided in the appendix.

Panel A: Prediction errors not included	(1)	(2)	(3)	(4)
Estimation Method	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
In-Sample R^2	0.66	-0.20	0.69	0.73
10 fold CV R^2	0.06	-3.27	0.11	0.18
Out-of-Sample R^2	0.04	-3496.42	0.11	0.15
Panel B: Prediction errors included as a factor	**(1)**	**(2)**	**(3)**	**(4)**
Estimation Method	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
In-Sample R^2	0.66	-0.26	0.70	0.73
10 fold CV R^2	0.08	-3.49	0.12	0.18
Out-of-Sample R^2	0.07	-0.50	0.11	0.14

sample R^2 values increase markedly. In columns (3) and (4) there is no change to the out-of-sample R^2 values.

The results in columns (3) and (4) may seem at odds with the Panel B results from 11. How can both hold at the same time? Table 11 is about linear predictability. Table 12 columns (3) and (4) are about forest based predictability. These are not the same thing.

In the machine learning literature is often found that ensembles of algorithms outperform the individual algorithms, see Hastie et al. (2009). Usually the ensemble is formed by voting across algorithms. Here we are only considering two approaches Fama-MacBeth regressions and gradient boosting. Combining the Tables we see that a two step procedure will improve the forecasts. Step 1 is to use the gradient boosting. Step 2 takes the result from the gradient boosting and then uses it in a linear regression. Table 11 shows that this second step will improve predictions relative to just using the gradient boosting alone.

How do we interpret the results from Tables 11 and 12? It is clear the algorithms reject the Bordalo et al. (2021) interpretation of rational expectations. The algorithms generate quite similar patterns of estimated coefficients to those from stock analysts. So the Nagel (2021) as-if interpretation seems reasonable. On the other hand the algorithms treat all observation equally at the start, and they have no emotions in the usual sense of the term. The negative coefficients may have a different source than has been suggested in the literature.

This section shows that both Fama-MacBeth and gradient boosting produce profit prediction mistakes that are linearly predictable. Improvements in the profit prediction are likely obtainable using ensemble methods. Such an investigation would be interesting and might help cast light on issues like firm credit ratings. However, it is outside the scope of this paper.

10 Conclusion

This paper compares firm profit predictions based on Fama-MacBeth regressions to predictions based on gradient boosting. Gradient boosting provides higher quality predictions due to their ability to include many more factors. The predictions are evaluated directly and also in three test settings; one from behavioral finance, one from corporate finance, and one from asset pricing.

When test from behavioral finance are applied to the predictions, the predictions apparently 'overreact'. This is true both of predictions based on Fama-MacBeth and those based on gradient boosting. The algorithms generate predictions that are 'too optimistic' in good times and 'too pessimistic' in bad times. When testing human predictions, this pattern of estimated coefficients has been attributed to emotional human decision making (Bordalo et al., 2021). Of course, neither Fama-MacBeth nor gradient boosting are emotional in the usual sense of the term.

In corporate finance investors would like to fund profitable opportunities and avoid unprofitable investments. According to Frank and Sanati (forthcoming) actual profit and predicted profit affect firm financing decisions differently. As in that model, we find that when a firm expects future profits it tends to issue equity and reduce debt. When a firm has current profits it tends to repurchase equity and issue debt.

In asset pricing profits and/or profit expectations are commonly thought to affect the cross section of stock returns (Novy-Marx, 2013; Fama and French, 2015). The profit predictions from gradient boosting provide a potentially useful alternative proxy to income (Fama and French, 2006) or to gross profit (Novy-Marx, 2013), when studying the cross section of stock returns. The properties are generally similar but the magnitudes seem somewhat stronger.

In theory firms try to maximize expected profits. We find that many actual firm actions and market values are readily understood in that way. There may be room to further improve the profit predictions by exploiting ensembles, or perhaps by using deep learning. Whether such technical improvements will alter our understanding of the role of expected firm profits deserves future investigation.

References

Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar, "Overreaction and Working Memory," *NBER Working Paper*, 2020, (w27947).

Aharoni, Gil, Bruce Grundy, and Qi Zeng, "Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis," *Journal of Financial Economics*, 2013, *110* (2), 347–357.

Ai, Hengjie, Murray Z Frank, and Ali Sanati, "The Trade-Off Theory of Corporate Capital Structure," *Oxford Research Encyclopedia of Economics and Finance*, 2021, *https://ssrn.com/abstract=3595492.*

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri V Nikolaev, "Deflating profitability," *Journal of Financial Economics*, 2015, *117* (2), 225–248.

Bates, Stephen, Trevor Hastie, and Robert Tibshirani, "Cross-validation: what does it estimate and how well does it do it?," *https://arxiv.org/abs/2104.00673*, 2021.

Bordalo, Pedro, Nicola Gennaioli, Andrei Shleifer, and Stephen J Terry, "Real credit cycles," Technical Report, National Bureau of Economic Research, Paper 28416 2021.

Campbell, John Y, *Financial decisions and markets: a course in asset pricing*, Princeton University Press, 2017.

Cerda, Patricio and Gaël Varoquaux, "Encoding high-cardinality string categorical variables," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

Chen, Tianqi and Carlos Guestrin, "XGBoost: A scalable tree boosting system," in "Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining" ACM 2016, pp. 785–794.

D'Amour, Alexander, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D Hoffman et al., "Underspecification presents challenges for credibility in modern machine learning," *arXiv preprint arXiv:2011.03395*, 2020.

Danis, András, Daniel A Rettl, and Toni M Whited, "Refinancing, profitability, and capital structure," *Journal of Financial Economics*, 2014, *114* (3), 424–443.

Diebold, Francis X, "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests," *Journal of Business and Economic Statistics*, 2015, *33* (1), 1–1.

— **and Robert S Mariano**, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 2002, *20* (1), 134–144.

Eckbo, B Espen and Michael Kisser, "Tradeoff theory and leverage dynamics of high-frequency debt issuers," *Review of Finance*, 2021, *25* (2), 275–324.

Efron, Bradley and Trevor Hastie, *Computer age statistical inference*, Cambridge University Press, 2016.

Erel, Isil, Léa H Stern, Chenhao Tan, and Michael S Weisbach, "Selecting directors using machine learning," *Review of Financial Studies*, 2021, *34* (7), 3226–3264.

Fama, Eugene F and James D MacBeth, "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 1973, *81* (3), 607–636.

— **and Kenneth French**, "Dividend yields and expected stock returns," *Journal of Financial Economics*, 1988, *22* (1), 3–25.

— **and Kenneth R French**, "Testing trade-off and pecking order predictions about dividends and debt," *Review of Financial Studies*, 2002, *15* (1), 1–33.

— **and** — , "Profitability, investment and average returns," *Journal of Financial Economics*, 2006, *82* (3), 491–518.

— **and** — , "A five-factor asset pricing model," *Journal of Financial Economics*, 2015, *116* (1), 1–22.

Frank, Murray Z and Ali Sanati, "Financing corporate growth," *Review of Financial Studies*, forthcoming.

— **and Keer Yang**, "Does Finance Flow to High Productivity Firms?," *https://ssrn.com/abstract=3295140*, 2019.

— **and Vidhan K Goyal**, "Capital structure decisions: which factors are reliably important?," *Financial Management*, 2009, *38* (1), 1–37.

— **and** — , "The profits–leverage puzzle revisited," *Review of Finance*, 2015, *19* (4), 1415–1453.

Friedman, Jerome H, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, 2002, *38* (4), 367–378.

Green, Jeremiah, John RM Hand, and X Frank Zhang, "The characteristics that provide independent information about average us monthly stock returns," *Review of Financial Studies*, 2017, *30* (12), 4389–4436.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Empirical asset pricing via machine learning," *Review of Financial Studies*, 2020, *33* (5), 2223–2273.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, "… and the cross-section of expected returns," *Review of Financial Studies*, 2016, *29* (1), 5–68.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning: data mining, inference, and prediction, 2nd edition*, 2 ed., Springer Science & Business Media, 2009.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, "Which factors?," *Review of Finance*, 2019, *23* (1), 1–35.

Inoue, Atsushi and Lutz Kilian, "In-sample or out-of-sample tests of predictability: Which one should we use?," *Econometric Reviews*, 2005, *23* (4), 371–402.

Li, Bin and Alberto G Rossi, "Selecting Mutual Funds from the Stocks They Hold: A Machine Learning Approach," *Available at SSRN 3737667*, 2020.

Louppe, Gilles, Louis Wehenkel, Antonio Sutera, and Pierre Geurts, "Understanding variable importances in forests of randomized trees," *Advances in neural information processing systems*, 2013, *26*, 431–439.

Mitton, Todd, "Methodological variation in empirical corporate finance," *Review of Financial Studies*, forthcoming, *Available at SSRN 3304875.*

Myers, Stewart C, "The capital structure puzzle," *Journal of Finance*, 1984, *39* (3), 574–592.

Nagel, Stefan, *Machine Learning in Asset Pricing*, Princeton University Press, 2021.

Novy-Marx, Robert, "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 2013, *108* (1), 1–28.

Petersen, Mitchell A, "Estimating standard errors in finance panel data sets: Comparing approaches," *Review of Financial Studies*, 2009, *22* (1), 435–480.

van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira, "Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases," *NBER Working Paper*, 2020, (w27843).

Wang, Huijun and Jianfeng Yu, "Dissecting the profitability premium," in "AFA 2013 San Diego meetings paper" 2013.

Welch, Ivo and Amit Goyal, "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 2008, *21* (4), 1455–1508.

A Appendix: data sources in detail

Table A1: Variable Definitions

(a)

This table defines profitability measures and variables as in Fama and French (2006). The variables are constructed using Compustat Annual data. Time subscription t is omitted if the variable is measured contemporaneously. $\mathbb{1}$ is the indicator function.

Profitability Measures		
Y_t^G	gross profits	$SALE_t - COGS_t$
π_t^G	gross profitability	$(SALE_t - COGS_t)/AT_t$
π_t^{GLag}	gross profitability (lag)	$(SALE_t - COGS_t)/AT_{t-1}$
Y_t^{OP}	operating profits	$SALE_t - COGS_t - XSGA_t$
π_t^{OP}	operating profitability	$(SALE_t - COGS_t - XSGA_t)/AT_t$
π_t^{OPLag}	operating profitability (lag)	$(SALE_t - COGS_t - XSGA_t)/AT_{t-1}$
Y_t^{Inc}	income profits	IB_t
π_t^{Inc}	income profitability	IB_t/BE_t
π_t^{IncLag}	income profitability (lag)	IB_t/BE_{t-1}
Other Variables		
BE_t	book equity	$AT_t - LT_t + TXDITC_t - PSTKL_t$
AC_t	accruals	$(ACT - CHE - LCT + DLC)_t -$ $-(ACT - CHE - LCT + DLC)_{t-1}$
DIV_t	dividends	$(DVPSX_F * CSHO)_t$

Variables from Fama and French (2006)		
lnB/M	log book to market	log(PRCC_F*CSHO)
lnMC	log market cap	log(PRCC_F*CSHO)
Neg Y	dummy variable for firms with negative profits	$\mathbb{1}(Y < 0)$
-AC/AT	negative accruals	$\min(AC, 0)/AT$
+AC/AT	positive accruals	$\max(AC, 0)/AT$
Δ_tA/A	asset growth	$AT_t/AT_{t-1} - 1$
No DIV	dummy variable for firms with zero dividends	$\mathbb{1}(DIV = 0)$
DIV/BE	dividends	DIV/BE
-AC/BE	negative accruals	$\min(AC, 0)/BE$
+AC/BE	positive accruals	$\max(AC, 0)/BE$
Δ_tBE/BE	asset growth	$BE_t/BE_{t-1} - 1$
DIV/AT	dividends	DIV/AT

Table A1: Variable Definitions – Continued

(b)

This table defines financial flow variables as in Frank and Goyal (2009). The variables are constructed using Compustat Annual data. Time subscription t is omitted if the variable is measured contemporaneously. The table also defines Goyal-Welch macro variables from Welch and Goyal (2008), which is directly downloaded from Welch's website. The GDP data is from Federal Reserve Economic Data.

Financial Flows		
D Net Iss	Debt Net Issuance	$(DLTIS - DLTR + DLCCH)/AT$
D Iss	Debt Issuance	$(\max(DLTIS, 0) + \max(DLCCH, 0))/AT$
D Rep	Debt Repurchase	$(\max(DLTR, 0) - \min(DLCCH, 0))/AT$
E Net Iss	Equity Net Issuance	$(SSTK - PRSTKC)/AT$
E Iss	Equity Issuance	$(\max(SSTK, 0))/AT$
E Rep	Equity Repurchase	$\max(PRSTKC, 0)/AT$
Frank and Goyal (2009) Variables		
mktbk	Market to Book	$(AT + (PRCC\_F * CSHO) - SEQ - TXDB)/AT$
tang	Tangibility	$PPENT/AT$
asset	Assets	$\log(AT)$
div	Dividend	$\mathbb{1}(DIV \neq 0)$
tdm	Market Leverage	$(DLTT + DLC)/(AT + PRCC\_F * CSHO - SEQ - TXDB)$
tda	Book Leverage	$(DLTT + DLC)/(AT)$
Welch and Goyal (2008) Macro Variables		
macro_bm	Book to Market	
macro_tbl	Treasury bill rate	
macro_ntis	Net Equity Expansion	
macro_svar	Stock Variance	
macro_dp	Dividend price ratio	
macro_ep	Earning Price Ratio	
macro_tms	Term Spread	
macro_dfy	Default Yield Spread	

Table A1: Variable Definitions – Continued

(c)

This table defines financial statement variables which serves as potential profitability predictors based on Frank and Yang (2019). The variables are constructed using Compustat Annual data. Time subscription t is omitted if the variable is measured contemporaneously.

Financial statements variables		
sale_at	Sales	$SALE/AT$
cogs_at	Cost of Goods Sold	$COGS/AT$
xsga_at	Selling General and Administrative Expense	$XSGA/AT$
dp_at	Depreciation and Amortization	DP/AT
xint_at	Interest and Related Expense - Total	$XINT/AT$
nopi_at	Nonoperating Income (Expense)	$NOPI/AT$
spi_at	Special Items	SPI/AT
txt_at	Income Taxes	TXT/AT
mii_at	Minority Interest (Income Account)	MII/AT
dvp_at	Dividends - Preferred/Preference	DVP/AT
cstke_at	Common Stock Equivalents	$CSTKE/AT$
xido_at	Extraordinary Items and Discontinued Operations	$XIDO/AT$
che_at	Cash and Short-Term Investments	CHE/AT
rect_at	Receivables	$RECT/AT$
invt_at	Inventories	$INVT/AT$
aco_at	Other Current Assets	ACO/AT
act_at	Total Current Assets	ACT/AT
ivaeq_at	Investment and Advances - Equity	$IVAEQ/AT$
ivao_at	Investment and Advances/Other	$IVAO/AT$
intan_at	Intangible Assets	$INTAN/AT$
ao_at	Other Assets	AO/AT
dlc_at	Current Debt	DLC/AT
ap_at	Accounts Payable	AP/AT
txp_at	Income Taxes Payable	TXP/AT
lco_at	Other Current Liabilities	LCO/AT
lct_at	Current Liabilities	LCT/AT
dltt_at	Long-Term Debt	$DLTT/AT$
lo_at	Other Liabilities	LO/AT
txditc_at	Deferred Taxes and Investment Tax Credit	$TXDITC/AT$
mib_at	Minority Interest (Balance Sheet)	MIB/AT
lt_at	Total Liabilities	LT/AT
pstk_at	Total Preferred Stock	$PSTK/AT$
ceq_at	Total Common Stock	CEQ/AT
seq_at	Stockholders Equity	SEQ/AT
oancf_at	Operating Activities Net Cash Flow	$OANCF/AT$
ivncf_at	Investing Activities Net Cash Flow	$IVNCF/AT$
fincf_at	Financing Activities Net Cash Flow	$FINCF/AT$
chech_at	Cash and Cash Equivalents Change	$CHECH/AT$

Table A1: Variable Definitions – Continued

(d)

This table defines firm-level predictive characteristics which serves as potential profitability predictors as in Green et al. (2017) and Gu et al. (2020).

absacc	Absolute accruals	Bandyopadhyay, Huang & Wirjanto	2010, WP	Compustat	Annual
aeavol	Abnormal earnings announcement volume	Lerman, Livnat & Mendenhall	2007, WP	Compustat+CRSP	Quarterly
age	# years since rst Compustat coverage	Jiang, Lee & Zhang	2005, RAS	Compustat	Annual
baspread	Bid-ask spread	Amihud & Mendelson	1989, JF	CRSP	Monthly
beta	Beta	Fama & MacBeth	1973, JPE	CRSP	Monthly
betasq	Beta squared	Fama & MacBeth	1973, JPE	CRSP	Monthly
bm_ia	Industry-adjusted book to market	Asness, Porter & Stevens	2000, WP	Compustat+CRSP	Annual
cashdebt	Cash flow to debt	Ou & Penman	1989, JAE	Compustat	Annual
cashpr	Cash productivity	Chandrashekar & Rao	2009, WP	Compustat	Annual
cfp	Cash flow to price ratio	Desai, Rajgopal & Venkatachalam	2004, TAR	Compustat	Annual
cfp_ia	Industry-adjusted cash flow to price ratio	Asness, Porter & Stevens	2000, WP	Compustat	Annual
chatoia	Industry-adjusted change in asset turnover	Soliman	2008, TAR	Compustat	Annual
chcsho	Change in shares outstanding	Ponti & Woodgate	2008, JF	Compustat	Annual
chempia	Industry-adjusted change in profit margin	Soliman	2008, TAR	Compustat	Annual
chinv	Change in inventory	Thomas & Zhang	2002, RAS	Compustat	Annual
chmom	Change in 6-month momentum	Gettleman & Marks	2006, WP	CRSP	Monthly
chpmia	Industry-adjusted change in pro t margin	Soliman	2008,TAR	Compustat	Annual
chtx	Change in tax expense	Thomas & Zhang	2011, JAR	Compustat	Quarterly
cinvest	Corporate investment	Titman, Wei & Xie	2004, JFQA	Compustat	Quarterly
convind	Convertible debt indicator	Valta	2016, JFQA	Compustat	Annual
currat	Current ratio	Ou & Penman	1989, JAE	Compustat	Annual
depr	Depreciation / PP&E	Holthausen & Larcker	1992, JAE	Compustat	Annual
divi	Dividend initiation	Michaely, Thaler & Womack	1995, JF	Compustat	Annual
divo	Dividend omission	Michaely, Thaler & Womack	1995, JF	Compustat	Annual
dolvol	Dollar trading volume	Chordia, Subrahmanyam & Anshuman	2001, JFE	CRSP	Monthly
dy	Dividend to price	Litzenberger & Ramaswamy	1982, JF	Compustat	Annual
ear	Earnings announcement return	Kishore, Brandt, Santa-Clara & Venkatachalam	2008, WP	Compustat+CRSP	Quarterly
egr	Growth in common shareholder equity	Richardson, Sloan, Soliman & Tuna	2005, JAE	Compustat	Annual
grcapx	Growth in capital expenditures	Anderson & Garcia-Feijoo	2006, JF	Compustat	Annual
grltnoa	Growth in long term net operating assets	Fair eld, Whisenant & Yohn	2003, TAR	Compustat	Annual
herf	Industry sales concentration	Hou & Robinson	2006, JF	Compustat	Annual
hire	Employee growth rate	Bazdresch, Belo & Lin	2014, JPE	Compustat	Annual
idiovol	Idiosyncratic return volatility	Ali, Hwang & Trombley	2003, JFE	CRSP	Monthly
ill	Illiquidity	Amihud	2002, JFM	CRSP	Monthly
indmom	Industry momentum	Moskowitz & Grinblatt	1999, JF	CRSP	Monthly
invest	Capital expenditures and inventory	Chen & Zhang	2010, JF	Compustat	Annual
lgr	Growth in long-term debt	Richardson, Sloan, Soliman & Tuna	2005, JAE	Compustat	Annual
maxret	Maximum daily return	Bali, Cakici & Whitelaw	2011, JFE	CRSP	Monthly
mom12m	12-month momentum	Jegadeesh	1990, JF	CRSP	Monthly
mom1m	1-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
mom36m	36-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
mom6m	6-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly

Table A1: Variable Definitions – Continued

ms	Financial statement score	Mohanram	2005, RAS	Compustat	Quarterly
mve_ia	Industry-adjusted size	Asness, Porter & Stevens	2000, WP	Compustat	Annual
nincr	Number of earnings increases	Barth, Elliott & Finn	1999, JAR	Compustat	Quarterly
orgcap	Organizational capital	Eisfeldt & Papanikolaou	2013, JF	Compustat	Annual
pchcapx_ia	Industry adjusted % change in capital expenditures	Abarbanell & Bushee	1998, TAR	Compustat	Annual
pchcurrat	% change in current ratio	Ou & Penman	1989, JAE	Compustat	Annual
pchdepr	% change in depreciation	Holthausen & Larcker	1992, JAE	Compustat	Annual
pchgm_pchsale	% change in gross margin - % change in sales	Abarbanell & Bushee	1998, TAR	Compustat	Annual
pchquick	% change in quick ratio	Ou & Penman	1989, JAE	Compustat	Annual
pchsale_pchinvt	% change in sales - % change in inventory	Abarbanell & Bushee	1998, TAR	Compustat	Annual
pchsale_pchrect	% change in sales - % change in A/R	Abarbanell & Bushee	1998, TAR	Compustat	Annual
pchsale_pchxsga	% change in sales - % change in SG&A	Abarbanell & Bushee	1998, TAR	Compustat	Annual
pchsaleinv	% change sales-to-inventory	Ou & Penman	1989, JAE	Compustat	Annual
pricedelay	Price delay	Hou & Moskowitz	2005, RFS	CRSP	Monthly
ps	Financial statements score	Piotroski	2000, JAR	Compustat	Annual
quick	Quick ratio	Ou & Penman	1989, JAE	Compustat	Annual
rd	R&D increase	Eberhart, Maxwell & Siddique	2004, JF	Compustat	Annual
rd_mve	R&D to market capitalization	Guo, Lev & Shi	2006, JBFA	Compustat	Annual
rd_sale	R&D to sales	Guo, Lev & Shi	2006, JBFA	Compustat	Annual
realestate	Real estate holdings	Tuzel	2010, RFS	Compustat	Annual
retvol	Return volatility	Ang, Hodrick, Xing & Zhang	2006, JF	CRSP	Monthly
roaq	Return on assets	Balakrishnan, Bartov & Faurel	2010, JAE	Compustat	Quarterly
roavol	Earnings volatility	Francis, LaFond, Olsson & Schipper	2004, TAR	Compustat	Quarterly
roeq	Return on equity	Hou, Xue & Zhang	2015, RFS	Compustat	Quarterly
roic	Return on invested capital	Brown & Rowe	2007, WP	Compustat	Annual
rsup	Revenue surprise	Kama	2009, JBFA	Compustat	Quarterly
salecash	Sales to cash	Ou & Penman	1989, JAE	Compustat	Annual
saleinv	Sales to inventory	Ou & Penman	1989, JAE	Compustat	Annual
salerec	Sales to receivables	Ou & Penman	1989, JAE	Compustat	Annual
secured	Secured debt	Valta	2016, JFQA	Compustat	Annual
securedind	Secured debt indicator	Valta	2016, JFQA	Compustat	Annual
sgr	Sales growth	Lakonishok, Shleifer & Vishny	1994, JF	Compustat	Annual
sin	Sin stocks	Hong & Kacperczyk	2009, JFE	Compustat	Annual
sp	Sales to price	Barbee, Mukherji, & Raines	1996, FAJ	Compustat	Annual
std_dolvol	Volatility of liquidity (dollar trading volume)	Chordia, Subrahmanyam & Anshuman	2001, JFE	CRSP	Monthly
std_turn	Volatility of liquidity (share turnover)	Chordia, Subrahmanyam, &Anshuman	2001, JFE	CRSP	Monthly
stdacc	Accrual volatility	Bandyopadhyay, Huang & Wirjanto	2010, WP	Compustat	Quarterly
stdcf	Cash flow volatility	Huang	2009, JEF	Compustat	Quarterly
tb	Tax income to book income	Lev & Nissim	2004, TAR	Compustat	Annual
turn	Share turnover	Datar, Naik & Radcliffe	1998, JFM	CRSP	Monthly
zerotrade	Zero trading days	Liu	2006, JFE	CRSP	Monthly

B Internet Appendix: Additional Tables

Table C1: Predicting profits without Profit$_t$

Profits are predicted for 1976 to 2015. Rolling estimation uses strictly prior data from 1964 to 2014 to make (pseudo) out-of-sample predictions. The variable being predicted is $\frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. The amount of variation explained is denoted R^2. Out-sample R^2 means that data from 1964 to 2014 is used to estimate the model on a rolling basis. The predictions are for 1976 to 2015. The models are fit using information from 1964 until time t to predict profits at time t+1. The out-of-sample R^2 is calculated as $1 - \frac{\sum(\hat{\pi}_{i,t+1}-\pi_{i,t+1})^2}{\sum(\pi_{i,t+1}-\pi_{i,t})^2}$. Estimation using the Fama and MacBeth (1973) method is denoted FM. Estimation using the Friedman (2002) method is denoted GBRT. When the data used as explanatory variables follows Fama and French (2006) except for Profit$_t$, it is denoted FF06 (no Profit$_t$). When the data used as explanatory variables are all factors used in Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), Gu et al. (2020), and except for Profit$_t$, it is denoted All (no Profit$_t$). When the data used as explanatory variables follows Fama and French (2006) and using year average Profit$_t$ instead of Profit$_t$, it is denoted FF06 (Average Profit$_t$). When the data used as explanatory variables are all factors used in Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), Gu et al. (2020), and using year average Profit$_t$ instead of Profit$_t$, it is denoted All (average Profit$_t$). Data construction details are provided in the appendix. Every cell in this table has 121,401 observations.

	(1)	(2)	(3)	(4)
Estimation Method	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
	(no Profit$_t$)	(no Profit$_t$)	(no Profit$_t$)	(no Profit$_t$)
Out-of-Sample R^2	-0.97	-3045.21	-0.44	0.08
Estimation Method	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
	(average Profit$_t$)	(average Profit$_t$)	(average Profit$_t$)	(average Profit$_t$)
Out-of-Sample R^2	-0.97	-3045.21	-0.41	0.09

Table C2: Predicting Gross profits and Income

Profits are predicted for 1976 to 2015. Rolling estimation uses strictly prior data from 1964 to 2014 to make (pseudo) out-of-sample predictions. The variable being predicted is $\frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. The amount of variation explained is denoted R^2. Out-sample R^2 means that data from 1964 to 2014 is used to estimate the model on a rolling basis. The predictions are for 1976 to 2015. The models are fit using information from 1964 until time t to predict profits at time t+1. The out-of-sample R^2 is calculated as $1 - \frac{\sum(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2}{\sum(\pi_{i,t+1} - \pi_{i,t})^2}$. Estimation using the Fama and MacBeth (1973) method is denoted FM. Estimation using the Friedman (2002) method is denoted GBRT. When the data used as explanatory variables follows Fama and French (2006) it is denoted FF06. When the data used as explanatory variables are all factors used in Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) it is denoted All. Gross profit means that profitability is measured as $\frac{\text{gross profit}(t+1)}{\text{total assets}(t+1)}$. Income means that profitability is measured as $\frac{\text{income}(t+1)}{\text{total assets}(t+1)}$. Data construction details are provided in the appendix. Every cell in this table has 121,401 observations.

	(1)	(2)	(3)	(4)
Estimation Methods	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
Profit Measure	Gross profit	Gross profit	Gross profit	Gross profit
Out-of-Sample $\{R^2\}$	0.04	-1864.29	0.06	0.11
Estimation Methods	FM	FM	GBRT	GBRT
Data	FF06	All	FF06	All
Profit Measure	Income	Income	Income	Income
Out-of-Sample $\{R^2\}$	0.09	-200.21	0.13	0.10

Table C3: What type of firms have more predictable profits?

This table reports several firm attributes for firms sorted into quintiles annually based on profits predictability. For each firm year, the predictability is calculated as negative squared error, $-(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2$. In columns (1)-(3), the prediction model is estimated following Fama and MacBeth (1973), with explanatory variables from Fama and French (2006). In columns (4) - (6), the prediction model is estimated following Friedman (2002), with the explanatory variables from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. The hyperparameters are set in recursive evaluation tuning method. In columns (7)-(9), firms sorted into quintiles annually, based on the predictability difference between GBRT, All Recursive model and FM, FF06 model. $H-L$ is the difference between the highest and lowest decile value. The t-stats in the brackets are for tests that the $H-L$ values are different from zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	GBRT, All Recursive			FM, FF06			GBRT, All Recursive - FM, FF06		
	Lowest	Highest	H - L	Lowest	Highest	H - L	Lowest	Highest	H - L
Negative SR_t+1	-493.74	-0.30	493.44***	-557.75	-0.31	557.44***	-94.47	156.09	250.55***
			(62.54)			(60.11)			(61.87)
Profit_t+1	-0.04	0.13	0.16***	-0.05	0.13	0.18***	0.08	-0.03	-0.11***
			(75.71)			(87.76)			(-48.86)
Profit_t	-0.00	0.13	0.13***	-0.01	0.13	0.14***	0.06	0.03	-0.03***
			(69.73)			(73.24)			(-12.54)
Total Assets_t	491.70	2944.42	2452.72***	508.02	2829.15	2321.13***	805.81	849.84	44.03
			(41.50)			(40.13)			(1.12)
R&D to sales_t	0.83	0.07	-0.76***	0.83	0.06	-0.77***	0.47	0.62	0.14***
			(-25.94)			(-26.62)			(4.52)
Market to Book_t	2.24	1.44	-0.80***	2.18	1.48	-0.70***	2.11	2.03	-0.08***
			(-51.72)			(-45.97)			(-4.53)
Cash Flow Volatility_t	13.75	1.94	-11.81***	13.70	1.76	-11.94***	8.85	11.07	2.21***
			(-18.84)			(-17.25)			(2.83)
Observations	24281	24280	48561	24281	24280	48561	24281	24280	48561

Table C4: What type of firms have more predictable profits?

This table reports several firm attributes for firms sorted into quintiles annually based on profits predictability. For each firm year, the predictability is calculated as negative squared error, $-(\hat{\pi}_{i,t+1} - \pi_{i,t+1})^2$. In columns (1)-(3), the prediction model is estimated following Fama and MacBeth (1973), with explanatory variables from Fama and French (2006). In columns (4) - (6), the prediction model is estimated following Friedman (2002), with the explanatory variables from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. The hyperparameters are set in cross-validation tuning method. In columns (7)-(9), firms sorted into quintiles annually, based on the predictability difference between GBRT, All CV model and FM, FF06 model. $H - L$ is the difference between the highest and lowest decile value. The t-stats in the brackets are for tests that the $H - L$ values are different from zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	GBRT, All CV			FM, FF06			GBRT, All CV - FM, FF06		
	Lowest	Highest	H - L	Lowest	Highest	H - L	Lowest	Highest	H - L
Negative $\text{SR}_t + 1$	-483.84	-0.29	483.55***	-557.75	-0.31	557.44***	-144.27	-431.35	-68.12***
			(62.58)			(60.11)			(-8.56)
$\text{Profit}_t + 1$	-0.04	0.13	0.16***	-0.05	0.13	0.18***	0.08	-0.04	-0.12***
			(76.22)			(87.76)			(-51.84)
Profit_t	-0.01	0.13	0.13***	-0.01	0.13	0.14***	0.05	0.03	-0.02***
			(71.39)			(73.24)			(-10.57)
Total Assets_t	485.19	2990.20	2505.01***	508.02	2829.15	2321.13***	664.19	770.96	106.77***
			(41.79)			(40.13)			(2.95)
R&D to sales_t	0.82	0.07	-0.75***	0.83	0.06	-0.77***	0.47	0.62	0.15***
			(-25.89)			(-26.62)			(4.74)
Market to Book_t	2.22	1.45	-0.76***	2.18	1.48	-0.70***	2.06	2.02	-0.04**
			(-50.00)			(-45.97)			(-2.07)
Cash Flow Volatility_t	13.31	1.65	-11.66***	13.70	1.76	-11.94***	8.47	11.59	3.12***
			(-21.22)			(-17.25)			(4.02)
Observations	24281	24280	48561	24281	24280	48561	24281	24280	48561

Table C5: Profits and financing flow decomposition: Recursive Tuned Hyperparameters

This table reports the effect of profitability and expected profitability on debt and and equity issuance and repurchasing. GBRT is used to forecast next period operating profit, π_{t+1}, using prior information available at time t including the variables from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. The hyperparameters are set in recursive evaluation tuning method. A constant term, year fixed effect, firm fixed effect, and firm control variables are included (Frank and Goyal, 2009). Standard errors are clustered at the firm level, and t-statistics are in parentheses. The dependent variable is always based on information that was available prior to the start of the time period. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every column in this table has $119,955$ observations.

	Debt$_t$			Equity$_t$			Assets Growth
Dep var	Net Iss	Gross Iss	Repur	Net Iss	Gross Iss	Repur	$\frac{\Delta_{t,t+1}AT}{AT_t}$
Profit$_t$	0.05***	-0.03***	-0.09***	-0.26***	-0.26***	0.00	-0.10**
	(5.88)	(-2.96)	(-7.61)	(-14.91)	(-14.80)	(0.30)	(-1.97)
E_tProfit$_{t+1}$	-0.01	0.04***	0.06***	0.10***	0.13***	0.03***	0.50***
	(-0.98)	(3.28)	(4.55)	(4.37)	(5.87)	(9.68)	(7.89)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.07	0.39	0.40	0.46	0.46	0.23	0.23

Table C6: Profits and financing flow decomposition: Cross-Validated Hyperparameters

This table reports the effect of profitability and expected profitability on debt and and equity issuance and repurchasing. GBRT is used to forecast next period operating profit, π_{t+1}, using prior information available at time t including the variables from Fama and French (2006), Frank and Goyal (2009), Frank and Yang (2019), and Gu et al. (2020) together. The hyperparameters are set in cross-validation tuning method. A constant term, year fixed effect, firm fixed effect, and firm control variables are included (Frank and Goyal, 2009). Standard errors are clustered at the firm level, and t-statistics are in parentheses. The dependent variable is always based on information that was available prior to the start of the time period. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every column in this table has $119,955$ observations.

	Debt$_t$			Equity$_t$			Assets Growth
Dep var	Net Iss	Gross Iss	Repur	Net Iss	Gross Iss	Repur	$\frac{\Delta_{t,t+1}AT}{AT_t}$
Profit$_t$	0.06***	-0.02*	-0.09***	-0.24***	-0.24***	0.00	-0.06
	(7.76)	(-1.85)	(-8.98)	(-13.67)	(-13.49)	(0.42)	(-1.20)
E_tProfit$_{t+1}$	-0.03***	0.03*	0.07***	0.06***	0.09***	0.03***	0.44***
	(-3.18)	(1.90)	(5.48)	(2.78)	(4.15)	(8.96)	(6.70)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.07	0.39	0.40	0.46	0.46	0.23	0.23

Table C7: Are prediction errors linearly predictable?

This follows Table 1 in Bordalo et al. (2021), regressing prediction error at $t+1$ on the information at time t. Prediction mistake at $t+1$ is Prediction Error$_{t+1} := p_{t+1} - F_t p_{t+1}$, where F_t is the model prediction based on date t data. Panel A uses GBRT and All variables with recursive tuning as the prediction model. Panel B uses GBRT and All variables with cross-validation tuning. Four different information available at time t are used. Prediction Error$_t := p_t - F_{t-1} p_t$ is prediction error at time t. Profit$_t = \frac{\text{operating profit}(t)}{\text{total assets}(t)}$ is profitability at time t. Investment$_t = \frac{\text{capital expenditure}(t)}{\text{total assets}(t)}$ is investment rate at time t. Debt issuance$_t$ is net debt issuance at time t. Column (1) has fewer observations due to unavailable forecasts when firm enters the sample at time t.

	(1)	(2)	(3)	(4)	(5)
Panel A	Prediction error$_{t+1}$				
Prediction Error$_t$	-0.31***				
	(-28.60)				
E$_t$ Profit$_{t+1}$		-0.12***			
		(-12.77)			
Profit$_t$			-0.18***		
			(-21.04)		
Investment$_t$				-0.03***	
				(-3.63)	
Debt issuance$_t$					0.04***
					(5.95)
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Method	GBRT, All	GBRT, All	GBRT, All	GBRT, All	GBRT, All
	Recursive	Recursive	Recursive	Recursive	Recursive
Observations	119955	106290	119955	119955	119955
Adjusted R^2	0.16	0.12	0.14	0.11	0.11
Panel B	Prediction error$_{t+1}$				
Prediction Error$_t$	-0.32***				
	(-29.42)				
E$_t$ Profit$_{t+1}$		-0.13***			
		(-14.22)			
Profit$_t$			-0.20***		
			(-23.23)		
Investment$_t$				-0.02***	
				(-2.95)	
Debt issuance$_t$					0.05***
					(6.59)
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Method	GBRT, All CV	GBRT, All CV	GBRT, All CV	GBRT, All CV	GBRT, All CV
Observations	119955	106290	119955	119955	119955
Adjusted R^2	0.16	0.11	0.14	0.10	0.10

Table C8: Excess returns and α sorted on expected profitability

This table shows monthly value-weighted average excess returns to portfolios sorted on different expected profitability measure, and α from time series regressions of these portfolios' return on the Fama French 3 factors for July 1976 to June 2016 sample period. Stocks from the July of calendar year $t+1$ to the June of calendar year $t+2$ are sorted into quintiles according to profit π_t or expected profit $E_t\pi_{t+1}$. π_t is profit in the fiscal year ending in calendar year t. $E_t\pi_{t+1}$ is expected profit in the fiscal year ending in calendar year $t+1$ predicted using information available in the fiscal year ending in calendar year t. In columns (2), we sort based on profit now (π_t). In columns (3), we sort based expected profit estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors, In columns (4), we sort based expected profit estimated using GBRT and all predictors. The expected profits are estimated similarly as in Table 2. We report the value-weighted portfolio excess returns of the portfolio with the lowest expected profit (Low) and the portfolio with the highest expected profit (High). The returns are in percentage points, so for example 2, means the monthly return is 2%. We calculate the spread between the highest and lowest expected profit portfolios (High - Low). For each portfilio, we also report the constant term (α) and the t-statistics (in the bracket) from time series regressions of regressing the portfolio returns on the Fama French 3 factors. Every portfolio has 480 monthly observations.

	(1)	(2)	(3)
Portfolio	Profit$_t$	E_tProfit$_{t+1}$	E_tProfit$_{t+1}$
	excess return		
Low	0.52	0.50	0.46
High	0.63	0.63	0.63
High - Low	0.10	0.13	0.16
	α		
Low	-0.21	-0.25	-0.29
	(-2.50)	(-2.46)	(-3.19)
High	0.18	0.18	0.18
	(3.42)	(3.60)	(3.62)
High - Low	0.39	0.43	0.47
	(3.42)	(3.38)	(3.96)
Method		FM, FF06	GBRT, All

Table C9: Profits and the cross section of expected returns

This table presents the monthly value-weighted average realized returns and spreads of portfolios formed on predicted returns. The predicted return are estimated monthly using Fama and MacBeth (1973) regressions for July 1976 to June 2016 sample period.

$$r_{i,t+1,k} = \theta_0 + \theta_1 \ln bkmkt_{i,t} + \theta_2 \ln mv_{i,t+1} + \theta_3 E_t \pi_{t+1}^{Lag} + \theta_1 \ln E_t \Delta_{t+1} AT/AT_t$$

$r_{i,t+1,k}$ is the return on stock i in the kth month of the 12 months from the July of calendar year $t+1$ to the June of calendar year $t+2$. i also denotes stock i in all the independent variables. Book-to-market, $\ln bkmkt_{i,t}$, is the logarithm of the book value of equity at the end of the fiscal year that ends in calendar year t divided by the market value of equity at the end of calendar year t. Size, $\ln mv_{i,t+1}$, is the logarithm of the market value of equity at the end of June of calendar year $t+1$. Expected profitability, $E_t\pi_{t+1}$, is the expected value of profit in the fiscal year ending in calendar year t+1. Profitability is operating profit scaled by lagged total assets, $\pi_{t+1}^{Lag} \equiv \frac{\text{operating profit}(t+1)}{\text{total assets}(t)}$. Expected asset growth, $E_t\Delta_{t+1}AT/AT_t$, is the expected growth of total assets in the fiscal year ending in calendar year t+1. In column (1) independent variables used to predict return are size and book-to-market only, which is the benchmark portfolio. In columns (2)-(4), we add expected investment and expected profitability. To be consistent, we measure expected investment estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors. In columns (2), we use profit now as expected profit measure ($E_t\pi_{t+1}^{Lag} \equiv \pi_t$). In columns (3), we measure expected profit estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors, In columns (4), we measure using GBRT and all predictors. The expected profits are estimated similarly as in Table 2. Predicted return from July of year t to June of year t+1, the fitted value from the Fama and MacBeth (1973) regression equation, are the product of average regressions slopes and explanatory variables at the end of June of year t. Stocks are sorted into quintiles according to their predicted return. Value-weighted average return is calculated for each group. We report the average realized returns of the portfolio with the lowest predicted return (Low) and the portfolio with the highest predicted return (High). The returns are in percentage points, so for example 2, means the monthly return is 2%. We calculate the spread between the highest and lowest predicted return portfolios (High - Low). We also compute the average difference and t-test statistics (in the brackets) between the "High - Low" spread in each column and the benchmark "High - Low" spread in column (1). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every portfolio has 480 monthly observations.

	(1)	(2)	(3)	(4)
Portfolio	size+value	Profit$_t$	E_tProfit$_{t+1}$	E_tProfit$_{t+1}$
Low	0.55	0.52	0.53	0.54
High	1.26	1.34	1.38	1.43
High - Low	0.70	0.82	0.85	0.90
Aver diff		0.12	0.15	0.19
t-statistic		1.68	2.19	3.12
Method			FM,FF06	GBRT,All

Table C10: Profits and the cross section of expected returns

This table presents the monthly value-weighted average realized returns and spreads of portfolios formed on predicted returns. The predicted return are estimated monthly using Fama and MacBeth (1973) regressions for July 1976 to June 2016 sample period.

$$r_{i,t+1,k} = \theta_0 + \theta_1 \ln bkmkt_{i,t} + \theta_2 \ln mv_{i,t+1} + \theta_3 E_t\pi_{t+1} + \theta_1 \ln E_t\Delta_{t+1}AT/AT_t$$

$r_{i,t+1,k}$ is the return on stock i in the kth month of the 12 months from the July of calendar year $t+1$ to the June of calendar year $t+2$. i also denotes stock i in all the independent variables. Book-to-market, $\ln bkmkt_{i,t}$, is the logarithm of the book value of equity at the end of the fiscal year that ends in calendar year t divided by the market value of equity at the end of calendar year t. Size, $\ln mv_{i,t+1}$, is the logarithm of the market value of equity at the end of June of calendar year $t+1$. $\text{Profit}_{t+1} = \frac{\text{operating profit}(t+1)}{\text{total assets}(t+1)}$. Expected profitability, $E_t\pi_{t+1}$, is the expected value of profit in the fiscal year ending in calendar year t+1. Expected asset growth, $E_t\Delta_{t+1}AT/AT_t$, is the expected growth of total assets in the fiscal year ending in calendar year t+1. In column (1) independent variables used to predict return are size and book-to-market only, which is the benchmark portfolio. In columns (2)-(4), we add expected investment and expected profitability. To be consistent, we measure expected investment estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors. In columns (2), we use profit now as expected profit measure ($E_t\pi_{t+1} \equiv \pi_t$). In columns (3), we measure expected profit estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors, In columns (4), we measure using GBRT and all predictors. The expected profits are estimated similarly as in Table 2. Predicted return from July of year t to June of year t+1, the fitted value from the Fama and MacBeth (1973) regression equation, are the product of average regressions slopes and explanatory variables at the end of June of year t. Stocks are sorted into 25 portfolios according to their predicted return. Value-weighted average return is calculated for each group. We report the average realized returns of the portfolio with the lowest predicted return (Low) and the portfolio with the highest predicted return (High). The returns are in percentage points, so for example 2, means the monthly return is 2%. We calculate the spread between the highest and lowest predicted return portfolios (High - Low). We also compute the average difference and t-test statistics (in the brackets) between the "High - Low" spread in each column and the benchmark "High - Low" spread in column (1). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every portfolio has 480 monthly observations.

	(1)	(2)	(3)	(4)
Portfolio	size+value	Profit_t	$E_t\text{Profit}_{t+1}$	$E_t\text{Profit}_{t+1}$
Low	0.50	0.36	0.43	0.33
High	1.08	1.45	1.52	1.52
High - Low	0.59	1.09	1.09	1.19
Aver diff		0.50	0.51	0.60
t-statistic		2.36	2.27	2.81
Method			FM,FF06	GBRT,All

Table C11: Profits and the cross section of expected returns: Gross Profit

This table presents the monthly value-weighted average realized returns and spreads of portfolios formed on predicted returns. The predicted return are estimated monthly using Fama and MacBeth (1973) regressions for July 1976 to June 2016 sample period.

$$r_{i,t+1,k} = \theta_0 + \theta_1 \ln bkmkt_{i,t} + \theta_2 \ln mv_{i,t+1} + \theta_3 E_t\pi_{t+1} + \theta_1 \ln E_t\Delta_{t+1}AT/AT_t$$

$r_{i,t+1,k}$ is the return on stock i in the kth month of the 12 months from the July of calendar year $t+1$ to the June of calendar year $t+2$. i also denotes stock i in all the independent variables. Book-to-market, $\ln bkmkt_{i,t}$, is the logarithm of the book value of equity at the end of the fiscal year that ends in calendar year t divided by the market value of equity at the end of calendar year t. Size, $\ln mv_{i,t+1}$, is the logarithm of the market value of equity at the end of June of calendar year $t+1$. Expected profitability, $E_t\pi_{t+1}$, is the expected value of profit in the fiscal year ending in calendar year t+1. Profit$_{t+1} = \frac{\text{gross profit}(t+1)}{\text{total assets}(t+1)}$. Expected asset growth, $E_t\Delta_{t+1}AT/AT_t$, is the expected growth of total assets in the fiscal year ending in calendar year t+1. In column (1) independent variables used to predict return are size and book-to-market only, which is the benchmark portfolio. In columns (2)-(4), we add expected investment and expected profitability. To be consistent, we measure expected investment estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors. In columns (2), we use profit now as expected profit measure ($E_t\pi_{t+1} \equiv \pi_t$). In columns (3), we measure expected profit estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors, In columns (4), we measure using GBRT and all predictors. The expected profits are estimated similarly as in Table 2. Predicted return from July of year t to June of year t+1, the fitted value from the Fama and MacBeth (1973) regression equation, are the product of average regressions slopes and explanatory variables at the end of June of year t. Stocks are sorted into 25 portfolios according to their predicted return. Value-weighted average return is calculated for each group. We report the average realized returns of the portfolio with the lowest predicted return (Low) and the portfolio with the highest predicted return (High). The returns are in percentage points, so for example 2, means the monthly return is 2%. We calculate the spread between the highest and lowest predicted return portfolios (High - Low). We also compute the average difference and t-test statistics (in the brackets) between the "High - Low" spread in each column and the benchmark "High - Low" spread in column (1). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every portfolio has 480 monthly observations.

	(1)	(2)	(3)	(4)
Portfolio	size+value	Profit$_t$	E_tProfit$_{t+1}$	E_tProfit$_{t+1}$
Low	0.50	0.44	0.45	0.45
High	1.08	1.36	1.36	1.34
High - Low	0.59	0.92	0.91	0.89
Aver diff		0.33	0.33	0.31
t-statistic		1.48	1.44	1.33
Method			FM,FF06	GBRT,All

Table C12: Profits and the cross section of expected returns: Income

This table presents the monthly value-weighted average realized returns and spreads of portfolios formed on predicted returns. The predicted return are estimated monthly using Fama and MacBeth (1973) regressions for July 1976 to June 2016 sample period.

$$r_{i,t+1,k} = \theta_0 + \theta_1 \ln bkmkt_{i,t} + \theta_2 \ln mv_{i,t+1} + \theta_3 E_t \pi_{t+1} + \theta_1 \ln E_t \Delta_{t+1} AT/AT_t$$

$r_{i,t+1,k}$ is the return on stock i in the kth month of the 12 months from the July of calendar year $t+1$ to the June of calendar year $t+2$. i also denotes stock i in all the independent variables. Book-to-market, $\ln bkmkt_{i,t}$, is the logarithm of the book value of equity at the end of the fiscal year that ends in calendar year t divided by the market value of equity at the end of calendar year t. Size, $\ln mv_{i,t+1}$, is the logarithm of the market value of equity at the end of June of calendar year $t+1$. Expected profitability, $E_t \pi_{t+1}$, is the expected value of profit in the fiscal year ending in calendar year t+1. Profit$_{t+1} = \frac{\text{income}(t+1)}{\text{total assets}(t+1)}$. Expected asset growth, $E_t \Delta_{t+1} AT/AT_t$, is the expected growth of total assets in the fiscal year ending in calendar year t+1. In column (1) independent variables used to predict return are size and book-to-market only, which is the benchmark portfolio. In columns (2)-(4), we add expected investment and expected profitability. To be consistent, we measure expected investment estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors. In columns (2), we use profit now as expected profit measure ($E_t \pi_{t+1} \equiv \pi_t$). In columns (3), we measure expected profit estimated using Fama and MacBeth (1973) regressions and Fama and French (2006) predictors, In columns (4), we measure using GBRT and all predictors. The expected profits are estimated similarly as in Table 2. Predicted return from July of year t to June of year t+1, the fitted value from the Fama and MacBeth (1973) regression equation, are the product of average regressions slopes and explanatory variables at the end of June of year t. Stocks are sorted into 25 portfolios according to their predicted return. Value-weighted average return is calculated for each group. We report the average realized returns of the portfolio with the lowest predicted return (Low) and the portfolio with the highest predicted return (High). The returns are in percentage points, so for example 2, means the monthly return is 2%. We calculate the spread between the highest and lowest predicted return portfolios (High - Low). We also compute the average difference and t-test statistics (in the brackets) between the "High - Low" spread in each column and the benchmark "High - Low" spread in column (1). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Every portfolio has 480 monthly observations.

	(1)	(2)	(3)	(4)
Portfolio	size+value	Profit$_t$	E_tProfit$_{t+1}$	E_tProfit$_{t+1}$
Low	0.50	0.46	0.48	0.38
High	1.08	1.25	1.31	1.35
High - Low	0.59	0.79	0.83	0.96
Aver diff		0.20	0.25	0.38
t-statistic		1.09	1.20	1.62
Method			FM,FF06	GBRT,All